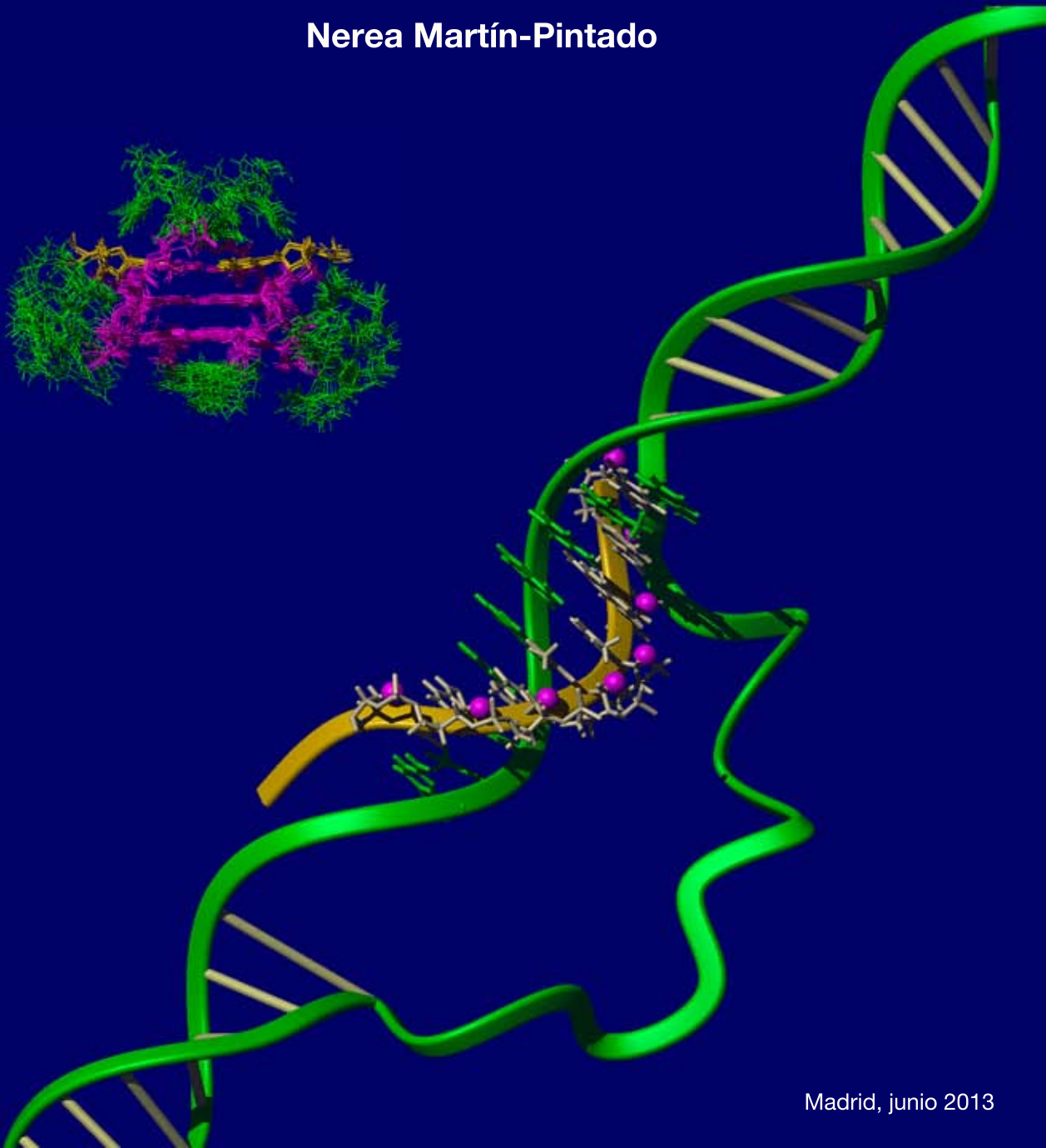


Estudio mediante RMN de la estructura y estabilidad de ácidos nucleicos con furanosas modificadas

Nerea Martín-Pintado



Tesis doctoral

**Estudio mediante RMN de la estructura y
estabilidad de ácidos nucleicos con furanosas
modificadas**

Memoria para optar al grado de Doctora por la Universidad Autónoma de Madrid
presentada por

Nerea Martín-Pintado Zugasti

Director

Carlos González Ibáñez

Instituto de Química Física "Rocasolano"
Grupo de Espectroscopía de RMN de Ácidos Nucleicos
Departamento Química Física Biológica
CSIC

Tutora

Carmen Navarro Ranninger

Programa de doctorado de Biofísica
Facultad de Química
Universidad Autónoma de Madrid
Madrid, mayo de 2013

A mis padres

AGRADECIMIENTOS

En primerísimo lugar dar las gracias a Carlos por su apoyo incondicional de estos años. Sin duda he tenido los mejores años de tesis que se podían tener gracias a ti. Después de tantos momentos en el despacho, congresos, “road trips” y demás eventos, va a ser muy duro irse a otra parte. Muchísimas gracias por todo.

A Masad Damha por acogerme en su laboratorio en mi primera estancia, y el haber sido para mí como un segundo “jefe” a lo largo de estos años.

Al resto de miembros del departamento, Manolo, Jorge, Marta, M^a Angeles, Doug, José Manuel, Cristina, Miguel Treviño, Irene, David y Flor, por vuestra ayuda, y a Alfredo especialmente por sus incontables historias y los momentos compartidos.

A todos esos compañeros que han pasado por el Roca o siguen en él. A Jorge Pedro, por toda su ayuda y por ser mi punto de referencia desde el momento que llegué al despacho. A Santi, porque el compartir años de tesis, cursos de doctorado, congresos y hasta vivir en el mismo edificio ha hecho que pasemos momentos geniales y que pueda contar contigo para cualquier cosa. A Miguel, por nuestros duros años juntos en el despacho, por nuestras múltiples apuestas jamás cumplidas, y porque incluso en los días más negros siempre eres capaz de decir alguna chorrada para hacerme reír.

A todos mis “Canadians” que me acogieron como una más de la familia. En especial a Maryam y Glen con los que he tenido el placer, no solo de trabajar, sino también de compartir un pedacito de sus vidas.

A mis amigas, esas con las que después de mil años sin vernos, el reencuentro es como si no hubiera un mañana.

A Roberto, por haberme apoyando en todo. Hasta en los peores momentos de estos años has estado cuando te necesitaba. Millones de gracias por ello y mucho más.

Finalmente agradecer a toda mi familia su amor incondicional. A mi hermano por todo su cariño, pero sobre todo por ir quitándome esas contracturas que la tesis me ha provocado. Y a mis padres, porque soy lo que soy por ellos, y porque jamás habría llegado hasta aquí sin su apoyo. GRACIAS.

ÍNDICE

1. Introducción y Objetivos	1
------------------------------------	----------

PARTE I: FUNDAMENTOS TEÓRICOS

2. Estructura de los ácidos nucleicos	9
--	----------

2.1. Estructura primaria de los ácidos nucleicos	11
---	-----------

- 2.1.1. Conformación del anillo de azúcar	12
- 2.1.2. Conformación del esqueleto azúcar fosfato	15
- 2.1.3. Conformación del ángulo glicosídico	15
- 2.1.4. Parámetros helicoidales: Posición y geometría de los pares de bases	17

2.2. Estructura secundaria de los ácidos nucleicos	19
---	-----------

- 2.2.1. Apareamiento entre bases: enlace de hidrógeno	20
- 2.2.2. Estructuras canónicas del DNA y RNA: Dúplex tipo A y tipo B	23
- 2.2.3. Estructuras no canónicas del DNA y RNA	28

2.3. Estabilización de las estructuras secundarias de los ácidos nucleicos	38
---	-----------

Bibliografía	40
---------------------	-----------

3. Nucleótidos modificados químicamente	45
--	-----------

3.1. Terapias antisentido y de RNA interferente: Conceptos básicos y mecanismos	47
--	-----------

- 3.1.1. Terapias antisentido y antígeno	48
- 3.1.2. Terapias basadas en la utilización de RNAs interferentes (siRNA)	51
- 3.1.3. Terapias basadas en la utilización de ribozimas	53

3.2. Nucleótidos modificados químicamente	53
--	-----------

- 3.2.1. Modificaciones de primera generación	54
- 3.2.2. Modificaciones de segunda generación	55
- 3.2.3. Modificaciones de tercera generación	56
- 3.2.4. Nucleótidos fluorados en la posición C2' del azúcar	57
- 3.2.5. Modificaciones en las bases	59

3.3. Estabilidad de los oligonucleótidos modificados	60
---	-----------

Bibliografía	61
---------------------	-----------

4. Determinación estructural de ácidos nucleicos por RMN	65
---	-----------

4.1. La RMN: herramienta fundamental para la determinación estructural de ácidos nucleicos	67
---	-----------

- 4.1.1. Fundamentos físicos de la RMN	68
- 4.1.2. RMN bidimensional	69
- 4.1.3. Información estructural extraída de los experimentos de RMN	76

4.2. Cálculo de estructuras en disolución	83
--	-----------

- 4.2.1. Métodos de simulación	84
- 4.2.2. Campo de fuerzas	87
- 4.2.3. Dinámica molecular restringida	89
- 4.2.4. Parametrización	90

Bibliografía	91
---------------------	-----------

PARTE II: RESULTS

5. The solution structure of double helical arabino-nucleic acids (ANA and 2'F-ANA): effect of arabinoses in duplex-hairpin interconversion	95
Resumen	96
Abstract	97
5.1. Introduction	98
5.2. Results	99
- 5.2.1. Effect of ANA and 2'F-ANA modifications on duplex stability	99
- 5.2.2. Duplex-hairpin equilibrium	100
- 5.2.3. NMR assignments	103
- 5.2.4. Experimental constraints and structure calculations	105
- 5.2.5. Description of the duplex structure	107
- 5.2.6. Hairpin structure	108
5.3. Discussion	110
- 5.3.1. Comparison of ANA and 2'F-ANA with unmodified DNA structure	110
- 5.3.2. ANA stabilizes stem-loop structures	112
- 5.3.3. ANA vs 2'F-ANA duplexes-why such a difference in stability	113
- 5.3.4. Alternative genetic systems based on arabinose?	114
5.4. Conclusion	115
Bibliography	115
6. The solution structure of 2'F-ANA:RNA and ANA:RNA hybrid duplexes. Roles of structure and hydrogen bonding in their different stability	119
Resumen	120
Abstract	121
6.1. Introduction	122
6.2. Results	123
- 6.2.1. Duplex formation and thermodynamics	123
- 6.2.2. NMR assignment	125
- 6.2.3. Experimental constraints and structure calculations	127
- 6.2.4. Description of the structures	130
- 6.2.5. Fluorine chemical shifts	132
6.3. Discussion	133
- 6.3.1. Structural differences	133
- 6.3.2. Pseudohydrogen bonding and 2'X-nucleobase interactions	134
- 6.3.3. FC-H2''..O backbone interactions	137
6.4. Conclusion	137
Bibliography	138
7. Stabilizing FC-H..O electrostatic interactions in 2'F-substituted Nucleic Acids	141
Resumen	142
Abstract	143
7.1. Introduction	144
7.2. Results and discussion	146
- 7.2.1. Duplex formation and thermal stability	146
- 7.2.2. Experimental NMR data and structure calculations	149
- 7.2.3. Structural properties of H and S duplexes	151
- 7.2.4. H and S duplexes are devoid of 2'F..H8 pseudohydrogen bonds	153

- 7.2.5. The effect of FC-H..O backbone interactions	154
- 7.2.6. FC-H2'..O backbone interactions in other C2'fluorinated structures	156
- 7.2.7. Are FC-H..O interactions real hydrogen bonds?	157
- 7.2.8. 2'F effect in duplex stability: Backbone hydrogen bonds vs enhanced base stacking and base pairing	159
7.3. Conclusion	160
Bibliography	161
8. Effect of incorporation of 2'-deoxy-2'2'-difluorodeoxycytidine (gemcitabine) on the structure and stability of DNA:RNA hybrid duplexes	163
Resumen	164
Abstract	165
8.1. Introduction	166
8.2. Results	167
- 8.2.1. Effect of dFdC modification in hybrid duplex stability	167
- 8.2.2. NMR assignments	169
- 8.2.3. Experimental constraints and structure calculations	171
- 8.2.4. Comparison of the structures	173
8.3. Discussion	175
- 8.3.1. Structural comparison between [GEM] and [CON]	175
- 8.3.2. 2'F..H8 stabilizing interactions	177
8.4. Conclusion	177
Bibliography	178
9. Effect of furanose C2'-substitution on structure and stability of G-quadruplex	181
Resumen	182
Abstract	183
9.1. Introduction	184
9.2. Results	186
- 9.2.1. G-quadruplex formation and thermodynamics	186
- 9.2.2. NMR assignment and experimental constraints	189
9.3. Discussion	194
9.4. Conclusion	197
Bibliography	197
Discusión general	201
Conclusiones	207
Estructuras determinadas experimentalmente	213
Annex I: Material and Methods	215
Annex II: Tables	223
Acrónimos y abreviaturas	255

Capítulo 1

Introducción y objetivos

1.1. Introducción

Después de 60 años de la publicación de la estructura del ácido desoxirribonucleico (DNA), todavía estamos lejos de comprender todas las estructuras que los ácidos nucleicos pueden adoptar. Normalmente, el DNA adopta una estructura en la que dos cadenas con secuencias complementarias se asocian en sentido antiparalelo, formando una hélice que gira en sentido dextrógiro. Esta estructura constituye el paradigma de la Biología Estructural y ha servido para explicar muchas de las propiedades y funciones del DNA como encargado de almacenar y transmitir la información genética. Durante décadas se han realizado multitud de estudios para entender los factores que afectan dicha estructura, tales como la secuencia concreta de nucleótidos, diferentes entornos físicos (pH, fuerza iónica, etc.), o la presencia de otras biomoléculas que interaccionan con el DNA, como proteínas y pequeños ligandos.

La estructura de la doble hélice dextrógira no es la única que puede adoptar el DNA. Las llamadas estructuras no-canónicas, diferentes de la doble hélice, han atraído mucha atención en los últimos años. En el capítulo 2 de esta tesis se hace una breve revisión de este tipo de estructuras. De particular importancia son los cuádruplex de guaninas, cuyo interés se ha disparado recientemente debido a su implicación en el mantenimiento de los telómeros, su prevalencia en secuencias promotoras y, sobre todo, a su reciente observación *in vivo*.

Los avances en métodos de síntesis química han tenido una gran influencia en los estudios estructurales. Muchos oligonucleótidos modificados representan lesiones que ocurren de manera natural en la célula. El estudio de estas lesiones, y su reconocimiento por la maquinaria de reparación celular, ha sido motivo de gran interés durante muchos años por su obvia implicación en procesos cancerígenos.

Si bien durante el siglo XX la fuerza motriz de la mayor parte de los estudios estructurales en ácidos nucleicos ha sido su importancia biológica, la situación ha cambiado notablemente en la última década. Los avances en Nanociencia y en Química Supramolecular han hecho de los ácidos nucleicos, y sus numerosos derivados, polímeros muy prometedores para desarrollar nuevos materiales y

dispositivos a escala nanométrica. Nuevamente, los avances en métodos de síntesis química corren en paralelo con los estudios biofísicos y estructurales, pues la posibilidad de modular las propiedades físico químicas de las cadenas de ácidos nucleicos abre multitud de potenciales aplicaciones.

De particular importancia es el reciente descubrimiento de que algunos de estos ácidos nucleicos modificados son capaces de almacenar y transmitir información genética como lo hacen los ácidos nucleicos naturales. Los estudios estructurales de estos nuevos polímeros, con capacidades similares a los de los ácidos nucleicos, es de gran interés en el nuevo y fascinante campo de la Biología Sintética.

Volviendo al interés biológico, quizá la motivación más importante para estudiar la estructura de los ácidos nucleicos sea su potencial aplicación en el desarrollo de nuevos fármacos. Existen muchos fármacos que actúan sobre los ácidos nucleicos y el descubrimiento y desarrollo de nuevos ligandos de ácidos nucleicos con valor terapéutico tiene un notable interés. No obstante, existe otra estrategia para el diseño de nuevos fármacos en la que el ácido nucleico no es la diana, sino el agente terapéutico en sí mismo. En las terapias llamadas antisentido o de RNA interferente, el principal objetivo es evitar la expresión de un gen determinado mediante la introducción de oligonucleótidos diseñados para unirse a la secuencia de DNA o de *mRNA* del gen cuya expresión se quiere bloquear (véase sección 3.1). En las terapias antisentido, por ejemplo, la adición de la cadena oligonucleotídica formará un dúplex híbrido DNA:RNA, mientras que en el caso de las terapias de RNA interferente se formará un dúplex RNA:RNA. Este tipo de estrategia terapéutica ya se utiliza en el tratamiento de algunos tipos de cáncer ocular.

Sin embargo, uno de los principales inconvenientes para la utilización generalizada de estas terapias, es la estabilidad de los oligonucleótidos *in vivo*. Tanto el DNA como el RNA son degradados por las nucleasas del organismo muy rápidamente. El mecanismo propuesto para mejorar la resistencia a nucleasas se basa en la introducción de modificaciones químicas en la cadena oligonucleotídica (véase sección 3.2).

Por otra parte, el continuo desarrollo y perfeccionamiento de las técnicas de determinación estructural, tales como la difracción de Rayos X y, especialmente, la Resonancia Magnética Nuclear (RMN) (véase capítulo 4), así como de los métodos computacionales, está teniendo un notable impacto en el campo de los ácidos nucleicos. La determinación estructural mediante estas técnicas está bien adaptada para el estudio de ácidos nucleicos relativamente pequeños, poco flexibles y estables. Sin embargo, hay que destacar que la síntesis e incorporación de nucleótidos modificados a cadenas oligonucleotídicas ha permitido ampliar estos estudios estructurales a biomoléculas que anteriormente eran inaccesibles para estas técnicas. Estos efectos son especialmente acusados al estudiar algunas formas no-canónicas, tales como hélices cuádruples, o procesos de reconocimiento molecular con péptidos u otros ligandos.

El trabajo llevado a cabo durante estos años, y que se presenta en esta memoria, se ha beneficiado del uso de técnicas de determinación estructural de alta resolución (RMN) disponibles en el Instituto de Química Física “Rocasolano” del CSIC. La capacidad de síntesis de ácidos nucleicos con modificaciones químicas de última generación, como son los análogos fluorados, han sido sintetizados en el grupo del Prof. Masad Damha de la Universidad McGill de Montreal y del Prof. Ramón Eritja del IQAC-CSIC de Barcelona.

1.2. Objetivos

El objetivo general de la tesis es determinar el efecto que tienen algunas modificaciones químicas en la posición C2' de la furanosa en la estructura y estabilidad de los ácidos nucleicos. Nos hemos centrado en modificaciones aparentemente menores, como son introducir átomos de fluor en C2' (2'F-ANA, 2'F-RNA y dFdC*) y sus correspondientes análogos hidroxilados (ANA y RNA*), pero que tienen un notable impacto en las propiedades del oligonucleótido resultante.

* Ver sección **Acrónimos**

Los ácidos nucleicos basados en arabinosas y fluoroarabinosas pueden replicarse con las polimerasas adecuadas, así como transcribirse a partir de un molde de DNA. Estas moléculas, por tanto, pueden almacenar y transmitir información genética como lo hacen el DNA y RNA en la naturaleza. Este hecho ha despertado el interés en la estructura del ácido arabinonucleico y de su fluoro-derivado. En el capítulo 5 nos centramos en el estudio de estas dos modificaciones (ANA y 2'F-ANA) en la estructura y estabilidad de la doble hélice. Estas dos modificaciones tienen efectos contrapuestos en la estabilidad. Así, los nucleótidos de 2'F-ANA forman estructuras de doble hélice muy estables, mientras los de ANA no tienen tendencia a hibridarse. En dicho capítulo nos planteamos elucidar las razones estructurales de estas diferencias en estabilidad y, para ello, se ha estudiado un dúplex quimérico con nucleótidos de 2'F-ANA en los extremos, y nucleótidos de ANA en la zona intermedia. Los nucleótidos 2'F-ANA confieren suficiente estabilidad al dúplex como para ser estudiado mediante técnicas de RMN.

Históricamente, la principal motivación para estudiar modificaciones en C2' ha sido su potencial aplicación biológica (véase sección 3.2.3 y 3.2.4). Por ello, es interesante conocer el efecto de estas modificaciones en los dúplex resultantes de su hibridación con cadenas de RNA. En el capítulo 6, se lleva a cabo el estudio estructural detallado de dúplex híbridos 2'F-ANA:RNA y ANA:RNA con el objeto de entender las razones estructurales de su drástica diferencia de estabilidad y su potencial aplicación en terapias antisentido.

En estudios previos, se ha corroborado que la inclusión de nucleótidos basados en fluororribosas (2'F-RNA) en secuencias de RNA provoca una fuerte estabilización de éstas, lo que hace del 2'F-RNA una modificación muy prometedora en terapias basadas en RNA interferente. Un efecto análogo en la estabilidad se observa con 2'F-ANA. Sin embargo, la inclusión de ambas modificaciones en la misma cadena oligonucleotídica provoca una desestabilización drástica de la estructura. En el capítulo 7 de esta tesis se ha estudiado la influencia de nucleótidos 2'F-ANA y 2'F-RNA en dúplex quiméricos. Para entender el tipo de interacciones que provocan este efecto, se han determinado las estructuras de dos dúplex quiméricos, uno de ellos con alternancia de 2'F-RNA y 2'F-ANA en la cadena

oligonucleotídica y otro con secuencias continuas de cada modificación. El objetivo es determinar las interacciones que provocan la desestabilización o la estabilización relativa de estos dúplex.

Tanto 2'-F-RNA como 2'-F-ANA provocan efectos estabilizantes cuyas causas estructurales han sido estudiadas a lo largo de esta tesis. Cabe preguntarse si la doble sustitución con flúor provoca efectos sinérgicos y aumenta aún más la estabilidad, o por el contrario, se eliminan algunas de las interacciones estabilizantes que se han observado en los casos anteriores. En el capítulo 8, se ha estudiado el efecto de introducir un nucleótido difluorado en la posición C2' (dFdC) en un dúplex híbrido DNA:RNA. Además del interés de estudiar el efecto de la doble sustitución en la estructura, este nucleótido modificado es actualmente un agente quimioterapéutico (gemcitabina) muy utilizado en terapias combinadas contra un gran número de tipos de cáncer.

A lo largo de nuestro trabajo, ha resultado evidente que el efecto de las modificaciones en la posición C2' de la furanosa tiene un gran efecto en la estructura y estabilidad de la doble hélice. Estos resultados, junto con el gran interés que despiertan hoy en día las estructuras no canónicas de los ácidos nucleicos, hace evidente la siguiente pregunta: ¿Cuál es el efecto de modificaciones en C2' en estructuras no canónicas?. En el capítulo 9 de esta tesis nos hemos planteado estudiar el efecto de estas modificaciones en la estructura y estabilidad de hélices cuádruples de guaninas (véase sección 2.2.3.5). Nos hemos centrado, para ello, en la secuencia telomérica humana, actualmente uno de los cuádruplex más estudiados en la literatura y de mayor relevancia biomédica.

Capítulo 2

Estructura de los ácidos nucleicos

2.1. Estructura primaria de los ácidos nucleicos

La unidad básica repetitiva del DNA (ácido desoxirribonucleico) y RNA (ácido ribonucleico) se denomina nucleótido. El nucleótido está compuesto por tres fragmentos moleculares, un azúcar, un fosfato y una base heterocíclica (Figura 1).



Figura 1: Estructura química de un nucleótido.

El heterociclo en posición C1' del azúcar (Figura 1) es una base nitrogenada plana y aromática unida covalentemente mediante un enlace glicosídico (χ). Las bases se clasifican en función de su estructura en púricas y pirimidínicas. Las bases púricas son adenina (A) y guanina (G), y las pirimidínicas son citosina (C), timina (T) y uracilo (U) (Figura 2).

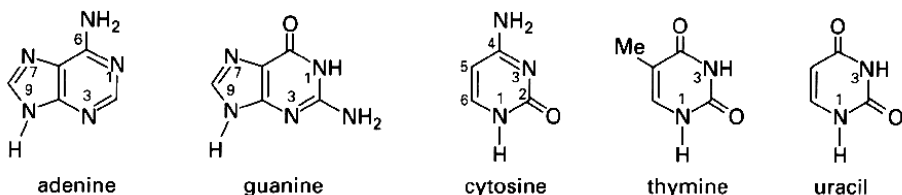


Figura 2: Estructura química de las bases nitrogenadas.

Por otro lado, hay dos tipos de azúcares que conforman los ácidos nucleicos, la β -D-ribose en el RNA y la β -D-2'-desoxirribose en el DNA. La unión entre la base y el azúcar se denomina nucleósido. La base uracilo sólo se encuentra en el RNA, mientras la timina se encuentra sólo en el DNA.

Los azúcares están fosforilados en la posición C5' (Figura 1). Los polinucleótidos se forman mediante la unión de diversos nucleótidos a través de un enlace

fosfodiéster entre el grupo hidroxilo 3' de un azúcar con el hidroxilo 5' del azúcar adyacente (Figura 3). A este enlace se le denomina fosfodiéster, y tiene una carga negativa que hace que el DNA y el RNA sean polianiones. Esta naturaleza polianiónica confiere a los ácidos nucleicos características de gran repercusión en su estructura secundaria y en su interacción con el medio que los rodea.

Los ácidos nucleicos, por tanto, son polianiones que forman largas cadenas con un esqueleto repetitivo a lo largo de toda la secuencia (fosfato-pentosa-fosfato) y una parte variable (base nitrogenada) (Figura 3). Las bases son las portadoras de la información genética, mientras que los grupos fosfato y azúcares tienen una misión estructural.

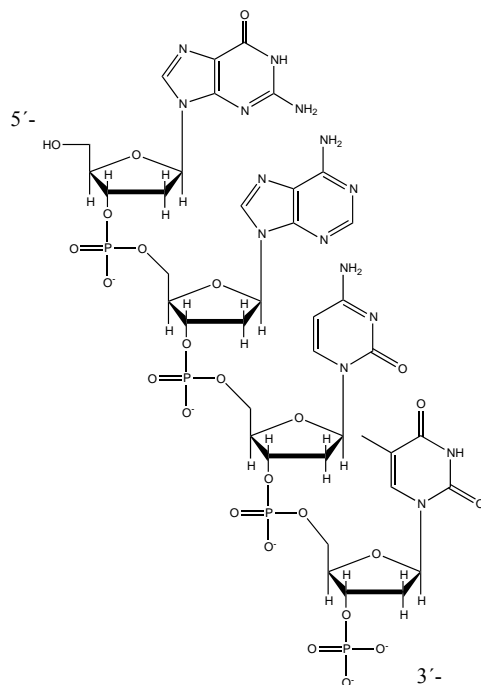


Figure 3: Cadena polinucleotídica de DNA con enlaces fosfodiéster entre nucleótidos.

2.1.1. Conformación del anillo de azúcar

Una de las características que le confiere flexibilidad estructural a los ácidos nucleicos es la conformación del anillo de ribosa y desoxirribosa. Los cinco átomos del anillo del azúcar adoptan normalmente una disposición fuera del plano (Figura

4) llamada “*puckering*”. Existen dos conformaciones principales: en forma “sobre” (E), o en forma “silla” (T)¹. La forma “sobre” tiene cuatro átomos en un plano y el quinto fuera de éste, y la forma “silla” presenta tres átomos en un plano y dos átomos situados a cada lado del mismo. Estas conformaciones a su vez se subdividen en la conformación *endo*, en la que los átomos que se encuentran fuera del plano están en el mismo lado del plano que C5', y en la conformación *exo* en la que dichos átomos están en el lado opuesto al átomo C5' (Figura 4).

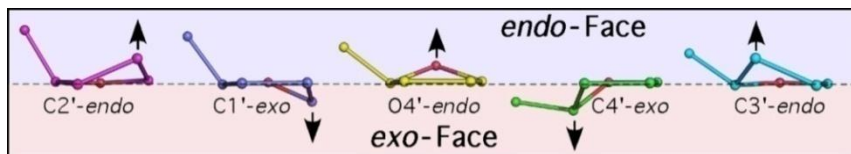


Figura 4: Conformaciones posibles del azúcar.

La conformación del anillo de pentosa se describe analíticamente introduciendo el concepto de pseudorrotación a través dos parámetros: el ángulo de fase de pseudorrotación (P), que indica el tipo de conformación o “*puckering*”, y la amplitud (v_m) que indica la magnitud máxima del mismo^{2,3}. Estos parámetros se obtienen a partir de las expresiones:

$$\tan P = \frac{(v_4 + v_1) - (v_3 + v_0)}{2v_2(\sin 36^\circ + \sin 72^\circ)} \quad v_m = \frac{v_2}{\cos P}$$

Así, la conformación del anillo de azúcar queda definida con sus cinco torsiones endocíclicas (v_0, v_1, v_2, v_3, v_4) (Figura 5):

$$v_0 = C4' - O4' - C1' - C2'$$

$$v_1 = O4' - C1' - C2' - C3'$$

$$v_2 = C1' - C2' - C3' - C4'$$

$$v_3 = C2' - C3' - C4' - O4'$$

$$v_4 = C3' - C4' - O4' - C1'$$

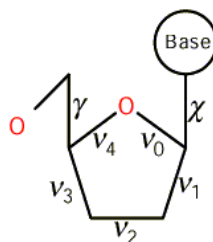


Figura 5: Nomenclatura y ángulos diedros del anillo de azúcar.

El ángulo de pseudorrotación² (P) en principio puede variar entre 0° y 360° , como se muestra en el ciclo de pseudorrotación del anillo de azúcar (Figura 6). Sin embargo, experimentalmente se ha comprobado que las desoxirribosas presentan dos conformaciones preferidas: la C3'-*endo* o norte (N) y la C2'-*endo* o sur (S). La conformación norte tiene ángulos de pseudorrotación entre 1° y 40° y la sur entre 130° y 180° ^{1,4}. Las conformaciones este (E) y oeste (W), en nucleótidos naturales, se corresponden con conformaciones menos favorecidas. Las ribosas normalmente adoptan una conformación C3'-*endo* (norte) mientras las desoxirribosas son más flexibles y favorecen un equilibrio entre norte y sur. La conformación del azúcar es determinante en la estructura global del ácido nucleico, ya que influye en la orientación de los átomos C1', C2' y C3'.

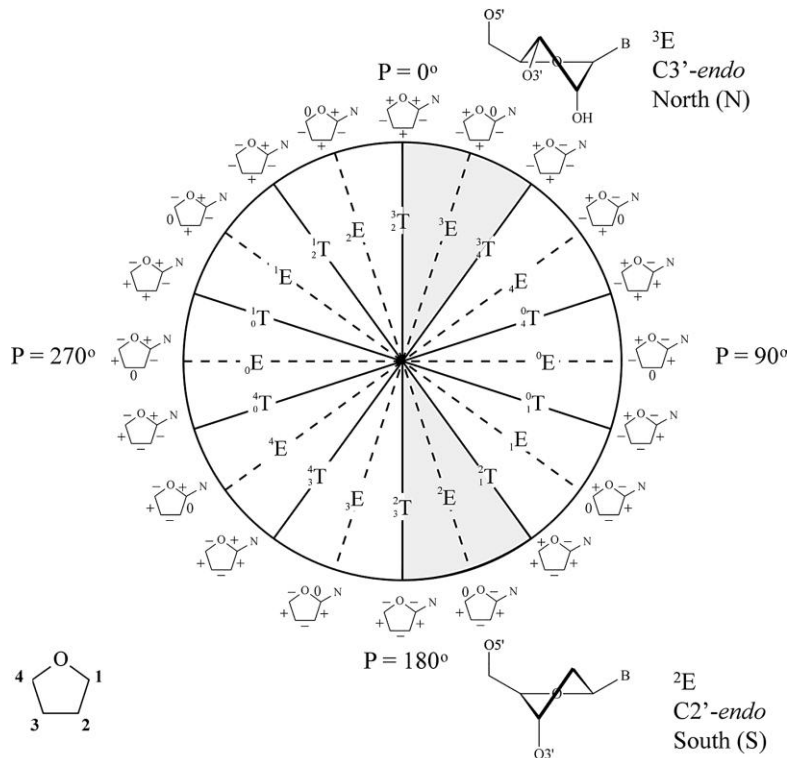


Figura 6: Ciclo de pseudorrotación del anillo de desoxirribosa, donde se muestra la relación entre la fase del ángulo de pseudorrotación (P) y las formas “sobre” (E) y “silla” (T). Las regiones en gris definen las conformaciones del azúcar más favorecidas en DNA y RNA.

2.1.2. Conformación del esqueleto azúcar-fosfato

El esqueleto azúcar-fosfato tiene una función estructural y está descrito por los seis ángulos de torsión α , β , γ , δ , ϵ , ζ :

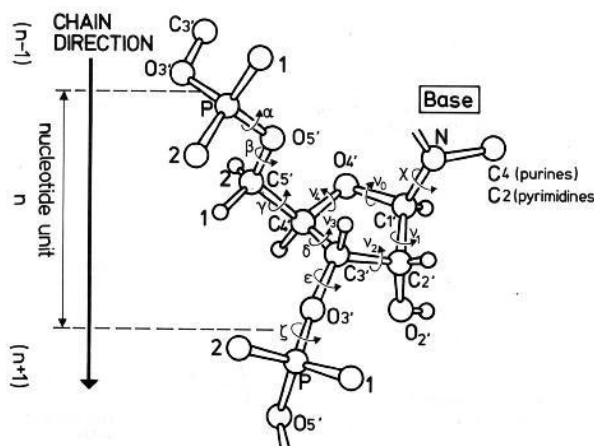


Figura 7: Ángulos de torsión del esqueleto en un nucleótido. Cada ángulo de rotación está indicado con una flecha.

Los ángulos representados en la Figura 7 no son independientes unos de otros, provocando que cambios conformacionales en el oligonucleótido, lleven asociados cambios concertados en varios de los ángulos de torsión, como veremos posteriormente.

2.1.3. Conformación del ángulo glicosídico

El ángulo glicosídico (χ) está definido por los átomos $O4'-C1'-N9-C4$ para las bases púricas y $O4'-C1'-N1-C2$ para las pirimidínicas. Este ángulo, en principio, podría tener un amplio rango de valores. Sin embargo, por razones estructurales, las bases nitrogenadas adoptan dos orientaciones principales en relación a la posición del azúcar, llamadas *anti* y *syn*^{5,6}. El ángulo en *anti* está entre -120° y -180° y en *syn* entre 0° y 90° . Debido a que la barrera energética es muy pequeña, en ciertas condiciones, se observa un equilibrio rápido entre ambas conformaciones (Figura 8).

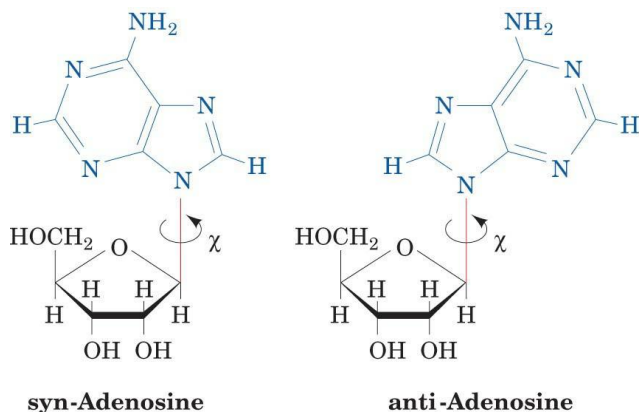


Figura 8: Conformación *syn* y *anti* en un nucleósido.

La conformación *anti* es aquella en la que el azúcar está en posición opuesta a la base y la *syn* aquella en la que el azúcar eclipsa la base (Figura 9)⁷. Normalmente la conformación más estable en ácidos nucleicos naturales es la *anti*, aunque en ciertos motivos estructurales (véase sección 2.2) también se observa la conformación *syn*. Normalmente existe cierta correlación entre el *puckering* del azúcar y el ángulo glicosídico, provocando que todos los ángulos que describen el nucleótido estén interrelacionados.

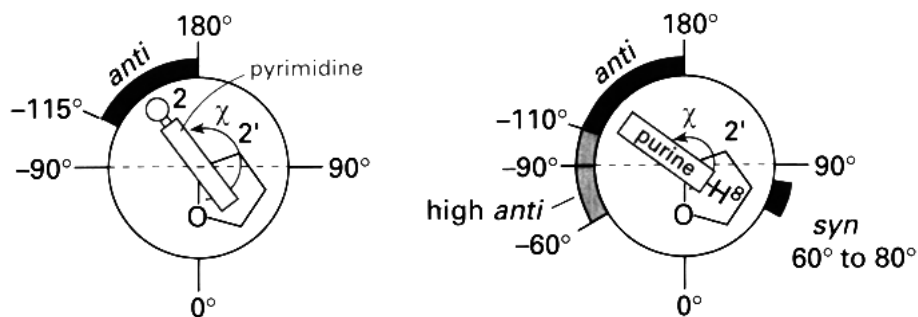


Figura 9: Conformación *anti* y *syn* de pirimidinas (izquierda) y purinas (derecha) respecto al azúcar. $\text{Syn}\chi = 0^\circ \pm 90^\circ$ y $\text{anti}\chi = 180^\circ \pm 90^\circ$.

2.1.4. Parámetros helicoidales: Posición y geometría de los pares de bases

Además de los ángulos que describen la conformación del nucleótido, también se utilizan los denominados parámetros helicoidales que son muy útiles para describir la posición y geometría de los pares de bases con respecto al eje helicoidal del dúplex, definido como eje z (Figura 10)⁴. Como describiremos en la siguiente sección, las bases de los ácidos nucleicos tienden a aparearse dando lugar a la asociación de cadenas distintas de polinucleótidos. Son necesarios seis parámetros para definir el par de bases, otros seis para definir la posición relativa de un par de bases con respecto al secuencial y cuatro para definir el par de bases con respecto al eje z .

Para el par de bases individual, los parámetros *y-displacement* (dy), *x-displacement* (dx), *inclination* (η) y *tip* definen la posición de cada par de bases con respecto al eje z o eje helicoidal. Los ejes de referencia x e y son perpendiculares a z y representan, respectivamente, la dirección que apunta hacia los lados del par de bases, y la que lo hace hacia el esqueleto de la cadena. Así, *x-* e *y-displacement* definen el desplazamiento de la base hacia los ejes x e y respectivamente (Figura 10). Por otro lado, los parámetros *inclination* y *tip* representan el ángulo entre el eje del par de bases y el plano xy (perpendicular al eje z). A la hora de definir la posición de una base con respecto a su complementaria, existen tres parámetros para definir las posiciones relativas (*stagger*, *stretch* y *shear*), y otras tres para los ángulos (*opening*, *propeller twist* y *buckle*) (Figura 10). *Shear* representa la posición relativa de las bases con respecto al eje x , *stretch* con respecto al eje y , y *stagger* determina la posición relativa de las bases en el plano xy . Con respecto a los ángulos, *opening* define la posición relativa de una base con respecto a su complementaria en el plano xy , *propeller twist* la rotación de una base con respecto a la complementaria en el par de bases y *buckle* la posición relativa de una base con respecto a la complementaria a lo largo del eje z .

Para definir las posiciones relativas entre dos pares de bases adyacentes en la hélice se definen los parámetros *rise*, *slide* y *shift*, y los ángulos relativos entre

dichos pares con los parámetros *twist*, *roll* y *tilt* (Figura 10). *Slide* mide la posición relativa de los pares de bases con respecto al eje *z*, *shift* con respecto al eje *x*, y *rise* con respecto al eje *y*. El parámetro *twist* define el ángulo entre pares de bases sucesivos basado en el cambio en la orientación del vector $C1'-C1'$. *Tilt* es la rotación de una base con respecto a la secuencial y con respecto al eje *x*, y *roll* mide el ángulo con respecto al eje del par de bases.

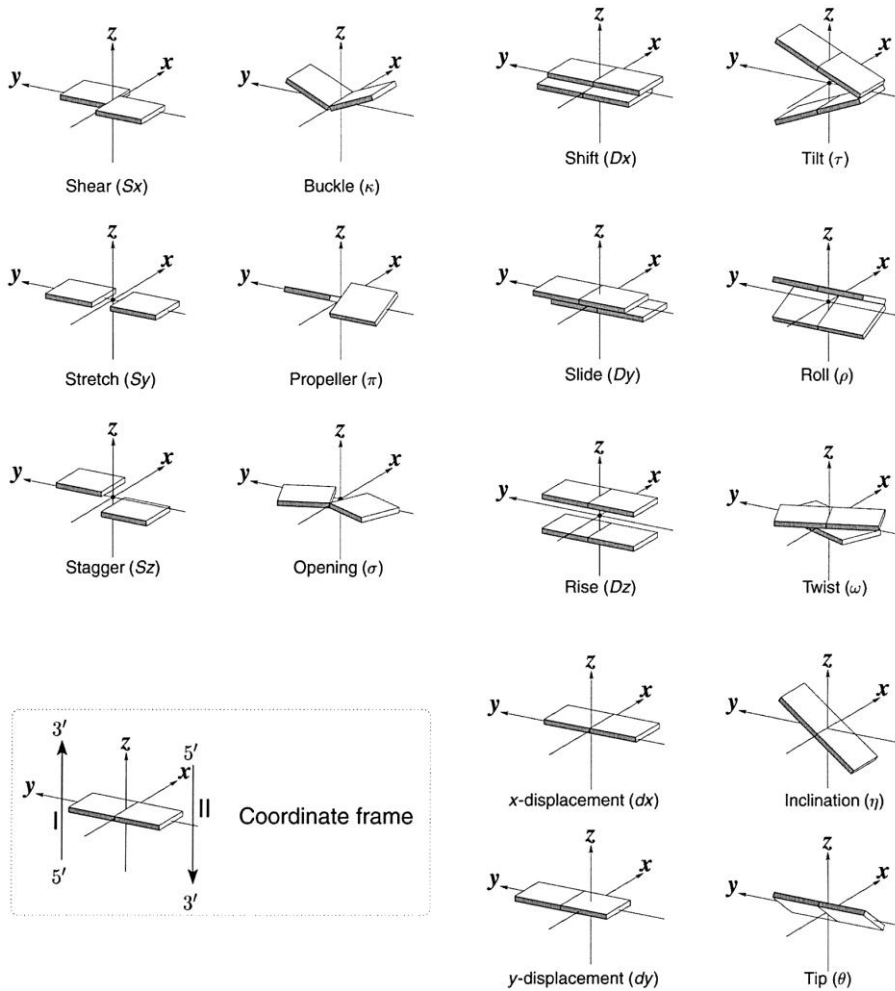


Figura 10: Esquema descriptivo de los parámetros helicoidales.

2.2 Estructura secundaria de los ácidos nucleicos

Las interacciones base-base, entre varias cadenas, permiten la formación de diferentes tipos de estructuras secundarias.

Las bases nitrogenadas se caracterizan por tener varias formas tautoméricas en función del pH del medio. El equilibrio se desplaza de la forma cetónica a la enólica en la guanina, timina y uracilo, y de la forma amino a la imino en la adenina y citosina. A pH fisiológico ($5 < \text{pH} < 9$), las formas cetónica o amino son las más favorecidas (Figura 11)⁸. El comportamiento ácido-base de los nucleótidos es fundamental a la hora de determinar su carga, estructura tautomérica y habilidad para ser aceptor o dador en los enlaces de hidrógeno entre las bases. En el rango fisiológico de pH, las bases nitrogenadas son neutras⁹, al igual que las pentosas, que tan solo en rangos de pH extremadamente básicos ($\text{pH} > 12$) pierden el protón de los grupos hidroxilo¹⁰.

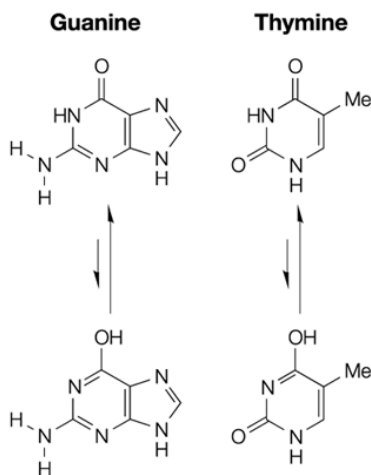


Figura 11: Formas tautoméricas de la guanina y timina. Equilibrio ceto-enol desplazado hacia la forma cetónica.

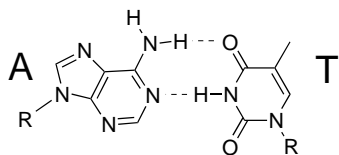
El grupo fosfato, por el contrario, a pH fisiológico posee carga negativa (-1), dando lugar a ácidos nucleicos polianiónicos. Las alteraciones en las bases y en los grupos fosfato por cambios en el pH, pueden provocar mutaciones y cambios en las estructuras secundarias que éstos conforman.

2.2.1. Apareamiento entre bases: enlace de hidrógeno

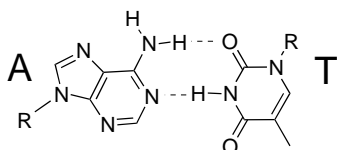
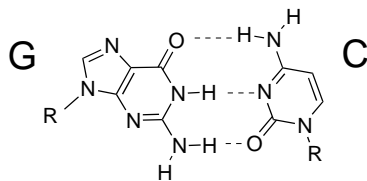
Las interacciones de hidrógeno son principalmente de carácter electroestático y juegan un papel fundamental en la estabilización de la estructura secundaria, tanto de proteínas como de ácidos nucleicos¹¹⁻¹³. El enlace de hidrógeno se forma si un átomo de hidrógeno (dador) se une a dos átomos X e Y de electronegatividad mayor, según el siguiente esquema:



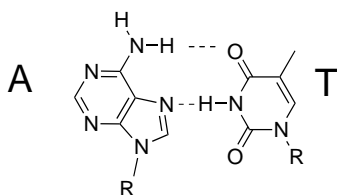
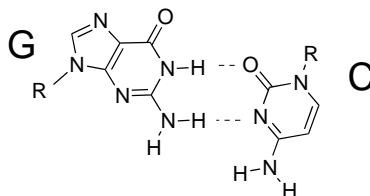
La fuerza del enlace de hidrógeno depende de las cargas localizadas en X, H e Y. Esta interacción está favorecida cuando el sistema de tres átomos es lineal, con una desviación no superior a 35° y una distancia entre los átomos entre 2.5 y 3.5 Å. Los enlaces de hidrógeno del tipo N-H \cdots N y N-H \cdots O son los responsables del reconocimiento entre bases de los oligonucleótidos (Figura 12). Los átomos de hidrógeno de los grupos N-H en adenina, guanina y citosina poseen carga positiva y son buenos dadores en los enlaces de hidrógeno. Por el contrario, las bases también actúan como aceptores a través de los átomos de nitrógeno y oxígeno (Figura 12).



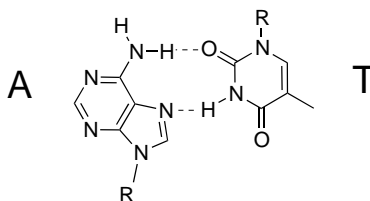
Watson - Crick



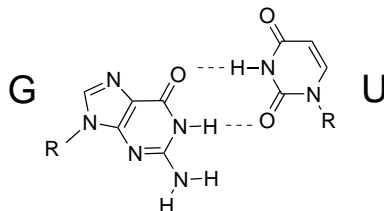
Reverse Watson - Crick



Hoogsteen



Reverse Hoogsteen



G-U Wobble

Figura 12: Apareamiento Watson-Crick y Watson-Crick reverso en los pares A·T y C·G. Apareamientos Hoogsteen y Hoogsteen reverso en pares A·T, y apareamiento Wobble en pares G·U.

Gracias a la complementariedad entre los grupos dadores y aceptores de las bases, se forman los pares de bases y consecuentemente las estructuras secundarias. La multitud de grupos dadores y aceptores en las bases permiten la formación de muchos tipos de pares de bases que darán lugar a distintos tipos de estructuras secundarias. La clasificación se divide en pares Watson-Crick (WC),

Hoogsteen (H), Watson-Crick reverso (rWC), Hoogsteen reverso (rH), *Wobble* y *mismatches*.

En el DNA, la forma más común de apareamiento entre pares de bases es la propuesta por Watson y Crick (**WC**) para la doble hélice¹⁴⁻¹⁶. Este tipo de apareamiento se caracteriza por la interacción de adenina (A) con timina (T) mediante dos enlaces de hidrógeno, y citosina (C) con guanina (G) por tres enlaces de hidrógeno entre cadenas antiparalelas (Figura 12). Independientemente del número de enlaces, hay coplanaridad entre las bases y las distancias en A·T y C·G son muy similares. Estas características dan lugar a la formación de estructuras helicoidales denominadas estructuras canónicas (ver sección 2.2.2.). El reconocimiento base-base mediante apareamiento Watson-Crick es clave en procesos de replicación, transcripción y traslación, por lo que el estudio detallado de las estructuras que conforman es de gran interés biológico.

El apareamiento tipo Hoogsteen (**H**)¹⁷, descubierto posteriormente en el DNA, se da en pares A·T y en pares G·C⁺ siempre y cuando la citosina esté protonada (Figura 12)¹⁸. Normalmente este tipo de apareamiento se produce en cadenas antiparalelas donde la base púrica (A o G) se encuentra en conformación *syn* (Figura 9). Este comportamiento se ha observado en cadenas compuestas únicamente por pares A·T, ácidos nucleicos modificados químicamente^{19,20}, hélices triples²¹, horquillas²², complejos DNA-proteína²³ y ocasionalmente en RNA²⁴. Recientemente se ha descubierto que, en estructuras secundarias del DNA, existe un equilibrio entre la forma Watson-Crick y la forma Hoogsteen²⁵, lo que pone de manifiesto la relevancia biológica que podría tener este tipo de pares de bases en los procesos moleculares.

El apareamiento tipo Hoogsteen también puede darse cuando el apareamiento entre hebras es paralelo²⁶. Sin embargo, normalmente, en el DNA paralelo la asociación entre hebras se favorece por pares Watson-Crick reverso (**rWC**)²⁷ (Figura 12). Se ha demostrado experimentalmente que una cadena de adeninas se puede unir a una de timinas (o uracilos en el RNA) para dar un dúplex paralelo con apareamiento rWC²⁸. En este tipo de estructuras la conformación mayoritaria de las bases es tipo *anti*.

Los apareamientos Hoogsteen reverso (**rH**) se han observado en estructuras de hélices triples (Figura 12). En concreto, se han encontrado en hélices triples tipo Pyr.Pur*Pur, donde las purinas de la tercera cadena se hibridan con las purinas del dúplex diana²⁹.

Los apareamientos tipo “*Wobble*” (**W**) son menos habituales ya que surgen del desplazamiento relativo de una base con respecto a la otra³⁰. El par G·U es muy frecuente en el RNA, donde se pierde la formación de un enlace de hidrógeno con respecto a los pares G·C de tipo Watson-Crick (Figura 12). Este tipo de apareamiento presenta dimensiones que imitan a la geometría adquirida por los pares de bases WC³¹.

Los “*mismatches*”, o apareamientos erróneos, se pueden clasificar en apareamientos purina:purina, pirimidina:pirimidina y purina:pirimidina³². La incorporación de este tipo de pares de bases es uno de los errores más comunes de los procesos de replicación y transcripción del DNA en la célula. Sin embargo, normalmente éstos son reparados con una eficiencia que depende en gran medida de la naturaleza del *mismatch* y de la secuencia que flanquea al mismo.

Finalmente mencionar que también se conocen apareamientos formados por bases modificadas. Estas bases surgen a partir de cambios en el entorno o bien son generadas por la célula durante el metabolismo. Por ejemplo, elementos cancerígenos como la luz ultravioleta (UV) o la radiación ionizante son capaces de inducir modificaciones en el DNA. Este efecto se produce, por ejemplo, en la alquilación de los grupos amino de adeninas.

2.2.2. Estructuras canónicas del DNA y RNA: Dúplex tipo A y tipo B

Las estructuras secundarias de los ácidos nucleicos se forman principalmente a partir de apareamientos entre las bases, y se pueden clasificar en dos grandes grupos:

- Estructuras canónicas: constan de dos cadenas asociadas de forma antiparalela para dar lugar a una estructura helicoidal dextrógira. Este tipo de estructura está basada en la descubierta por Watson y Crick en 1953 por difracción de rayos X¹⁴.

- Estructuras no canónicas: se dan mediante otros tipos de asociación entre dos o más cadenas polinucleotídicas.

Dado que en esta tesis se detallarán y discutirán los cambios conformacionales que las modificaciones químicas producen en algunas de estas estructuras, es muy importante conocer su topología.

Entre las estructuras canónicas más importantes están los dúplex tipo A y los dúplex tipo B, dobles hélices dextrógiras formadas por apareamiento de cadenas complementarias y antiparalelas³³. Los pares de bases se encuentran en el centro, perpendiculares al eje de la molécula, y apilados en su misma dirección. Exteriormente se encuentra el esqueleto azúcar-fosfato a lo largo de los lados de la hélice. Los pares de bases son tipo Watson y Crick (WC), uniendo cada base púrica o pirimidínica con su complementaria. Los pares A·T en DNA, o A·U en RNA, se mantienen unidos por dos enlaces de hidrógeno, mientras que en los pares C·G se forman tres enlaces. Esta propiedad de los pares de bases provoca que normalmente los pares A·T se abran o deformen más fácilmente que los pares C·G. La disposición, tanto de las bases como del esqueleto en estas dobles hélices, da lugar a dos surcos de distinta profundidad y anchura. Estos dos surcos se clasifican en surco mayor o “*major groove*” y surco menor o “*minor groove*”. El surco menor determina el surco del dúplex por la cara del átomo O2 en pirimidinas y el átomo N3 en purinas. El surco mayor se encuentra en la cara opuesta al surco menor (átomos N7 en purinas y C5 en pirimidinas). La anchura de los surcos se define como la mínima distancia entre grupos fosfato de cadenas complementarias menos el diámetro de van der Waals de un grupo fosfato (5.8 Å). Estos surcos normalmente tienen un papel fundamental en la interacción de la doble hélice con otras moléculas (proteínas o ligandos) y, por tanto, es un parámetro conformacional muy importante³⁴⁻³⁶.

Aunque estas características son comunes a las dobles hélices de DNA y RNA, los ácidos nucleicos pueden adoptar distintas conformaciones dependiendo de las condiciones del medio. Así, el DNA puede dar lugar a los denominados DNA tipo A y B mientras que el RNA adopta siempre la forma A.

2.2.2.1. Familia B

El DNA tipo B es la estructura clásica del DNA descrita por Watson y Crick, y la conformación más habitual del DNA en disolución¹⁵. Se caracteriza por una disposición *anti* para la torsión del ángulo glicosídico, un *puckering* del azúcar C2'-endo (sur) y una separación entre los pares de bases adyacentes de 3.4 Å. En total tiene 10 pares de bases por vuelta de hélice y un diámetro de hélice de 20 Å. Estas características dan lugar a una hélice con dos surcos (mayor y menor) de aproximadamente igual profundidad (8 Å), pero diferente anchura (Figura 13)³³. Mientras el surco menor tiene una anchura de 6 Å, el surco mayor es de aproximadamente 12 Å (Tabla 1). Los valores típicos de ángulos diedros se indican en la Tabla 2.

Tabla 1: Dimensiones de los surcos (en Å) en estructuras tipo A, B o Z.

	Surco mayor		Surco menor	
	Anchura	Profundidad	Anchura	Profundidad
A	2.2	13.0	11.1	2.6
B	11.6	8.5	6.0	8.2
Z	8.8	3.7	2.0	13.8

2.2.2.2. Familia A

La familia de dobles hélices de DNA tipo A se observa en cristales con condiciones de baja humedad y alta fuerza iónica. Por otro lado, el RNA siempre adopta este tipo de conformación, tanto en homoduplex como en híbridos RNA:DNA.

La doble hélice tipo A se caracteriza por ser más ancha, con 11 residuos por vuelta y un *puckering* del azúcar tipo C3'-endo (norte), diferente a la conformación C2'-endo adoptada en el DNA tipo B. Como hemos mencionado con anterioridad, los ángulos azúcar-fosfato no son independientes unos de otros. Esto provoca que la estructura global del DNA tipo A sea distinta a la del DNA tipo B. Los parámetros helicoidales (bases), ángulos de torsión y anchura y profundidad de los surcos son algunos de los afectados (Tabla 1 y 2). Así, los grupos fosfato consecutivos están más cerca que en la conformación B-canónica (5.9 Å vs 7 Å en el DNA tipo B). Entre los parámetros helicoidales se puede observar como la

inclinación, el *rise*, el *twist propeller* o el *x-displacement*, entre otros, sufren cambios. Mientras los valores de *rise* disminuyen, el *x-displacement* aumenta, provocando que las estructuras tipo A sean más achatadas y anchas, y que presenten el característico canal central (Figura 13). Los rangos preferidos de ángulos de torsión del esqueleto, desde α (P-O5') a ϵ (O3'-P) son similares a los encontrados en el DNA tipo B, excepto para el ángulo δ (Tabla 2)³³.

Tabla 2: Ángulos de torsión en (°) en los dúplex tipo A, B y Z.

	α	β	γ	δ	ϵ	ζ	χ
A	-52	175	42	79	-148	-75	-157
B	-30	136	31	143	-141	-161	-98
Z	-140	-137	51	95	-97	82	-154

Otra diferencia importante es la observada en la orientación del ángulo glicosídico C1'-N, que para DNA tipo A tiene un valor medio de -160° , mientras que en DNA tipo B tiene un valor de -100° . Además, son destacables las diferencias en los surcos. En este caso, el surco mayor es profundo (13 Å) y estrecho (2 Å) mientras el surco menor es más ancho (11 Å) y menos profundo (3 Å) (Figura 13). Este tipo de variaciones estructurales, consecuencia del polimorfismo de los ácidos nucleicos serán fundamentales a la hora de estabilizar estructuras compuestas por nucleótidos modificados químicamente.

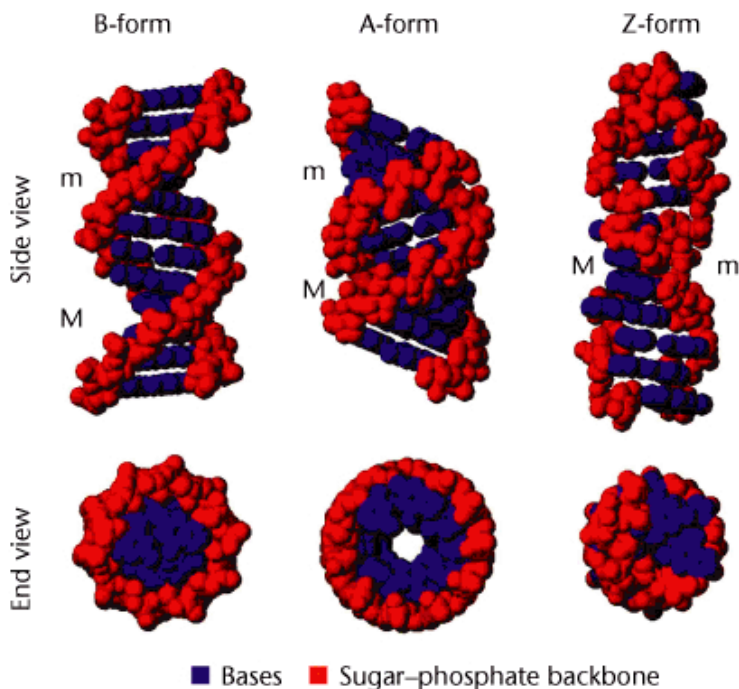


Figura 13: Representación esquemática de las dobles hélices A, B y Z. El surco mayor representado como M y el menor como m.

Es importante destacar que, aunque los valores dados son generales para estructuras tipo A y B canónicas, existen pequeñas variaciones en la estructura dependiendo de la secuencia estudiada. Por ejemplo, una secuencia con muchos pares C-G tendrá unos surcos y ángulos *twist* muy distintos a una secuencia con pares A-T. Uno de los ejemplos más estudiados, tanto por cristalografía como por Resonancia Magnética Nuclear (RMN), es el dodecámero de Dickerson-Drew d(CGCGAATTCGCG). La estructura de esta secuencia palindrómica de DNA fue determinada por primera vez en 1979 y se caracteriza porque contiene tanto pares C-G como A-T³⁷. Así, la región 5'-AATT muestra un estrechamiento del surco menor y unos valores *propeller twist* mucho mayores que los encontrados en la zona de pares C-G. El gran número de estudios realizados en esta estructura, así como la combinación de los dos tipos de pares de bases (C-G y A-T)³⁸⁻⁴⁰, han hecho que esta secuencia haya sido utilizada en los capítulos 5 y 7 para la determinación estructural de dobles hélices modificadas químicamente.

La forma predominante de un dúplex de DNA en disolución es la forma B, lo cual implica tener *a priori* cierta información estructural. Las condiciones óptimas para el estudio, tanto de RNA (siempre tipo A) como de DNA tipo B, mediante RMN, son las concentraciones de la muestra entre 0.5-3 mM y alta fuerza iónica (en nuestros estudios 25 mM fosfato sódico y 100 mM cloruro sódico). Para que se forme en disolución DNA tipo A es necesario que el dúplex esté en una disolución parcialmente deshidratada, como ocurre en mezclas de alcohol y agua.

Tabla 3: Comparación entre los tipos A, B y Z del DNA.

	Tipo de hélice		
	A	B	Z
Forma	La más ancha	Más estrecha	La más estrecha
Altura por par de bases	2.3 Å	3.4 Å	3.8 Å
Diámetro de la hélice	25.5 Å	23.7 Å	18.4 Å
Sentido de giro	Dextrógiro	Dextrógiro	Levógiro
Enlace glicosídico	<i>Anti</i>	<i>Anti</i>	<i>Anti</i> para C y T y <i>syn</i> para G
Par de bases/vuelta de hélice	11	10.4	12
Longitud de una vuelta de hélice	25.3 Å	35.4 Å	45.6 Å
Inclinación de los pares de bases respecto al plano del eje de la hélice	19°	1°	9°
Surco mayor	Estrecho y muy poco profundo	Ancho y bastante profundo	Plano
Surco menor	Muy amplio y poco profundo	Estrecho y bastante profundo	Muy estrecho y profundo

2.2.3. Estructuras no canónicas del DNA y RNA

Como ya hemos mencionado anteriormente, la gran cantidad de átomos dadores y aceptores de las bases nitrogenadas permite que se puedan combinar de distintas maneras y den lugar a estructuras muy diversas. Entre otras, se ha observado la formación de dúplex tipo Z, dúplex paralelos, horquillas, hélices tríples, hélices cuádruples, etc. Aunque comentaremos algunas de ellas brevemente, nos vamos a centrar en la estructura de las hélices cuádruples. El creciente interés de este

tipo de estructuras viene dado por su localización en zonas teloméricas y promotoras, que hacen de estas estructuras, dianas terapéuticas interesantes⁴¹.

2.2.3.1 Hélice Z levógira

Este tercer tipo de hélice de DNA fue descubierto por Rich *et al.* al resolver la estructura del hexámero d(CGCGCG)₂⁴². Posteriormente, en 1984, este tipo de estructura se confirmó también en RNA⁴³. En ambos casos, aun siendo un dúplex antiparalelo, la hélice es levógira (Figura 13). El esqueleto es irregular comparado con las estructuras anteriormente descritas, ya que los nucleótidos de citosina tienen una conformación C2'-*endo* (sur) y un ángulo glicosídico *anti*, mientras las guaninas tienen conformaciones C3'-*endo* (norte) y la base en *syn*. Esto provoca una distribución de los grupos fosfato en “zig-zag” que dio lugar al nombre de DNA tipo Z⁴⁴. En esta estructura los fosfatos se encuentran muy próximos entre sí, y en consecuencia, este tipo de dúplex se observa en condiciones de alta concentración de sales (como MgCl₂, NaCl, NaClO₄), o de alcoholes (etanol o perclorato), capaces de reducir la repulsión electrostática entre estos grupos fosfato. También parece necesario que sea una secuencia tipo poli(dG-dC). Los surcos de DNA tipo Z se diferencian mucho de las formas A y B (Tabla 1). Las diferencias con respecto a ángulos de torsión, comparado con los dúplex A y B, también se pueden observar en la Tabla 2.

2.2.3.2. Horquillas o “hairpins”

Las horquillas constan de un fragmento de doble hélice (“*stem*”) y un giro de cadena sencilla (“*loop*”). Para la conexión de los dos lados del *stem* es necesaria una inversión de la dirección del esqueleto del oligonucleótido. Este efecto viene dado por el *loop* que permite el cambio de sentido de la hebra y la hibridación antiparalela en el *stem* (Figura 14). Este tipo de estructuras se han observado principalmente en el RNA, aunque también se han encontrado ciertas secuencias de DNA que son capaces de formar estructuras tipo horquilla o “*hairpin*” para dar lugar a estructuras cruciformes⁴⁵. Este tipo de secuencias palindrómicas se encuentran próximas a lugares de unión y regulación génica y en zonas promotoras, lo que sugiere una posible implicación biológica. Además, este tipo de estructura en horquilla muestra mayor resistencia a la degradación por

exonucleasas, lo que ha despertado también su interés como posible fármaco antisentido (ver capítulo 3).

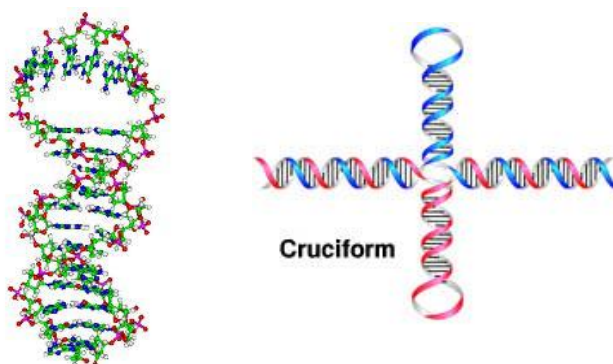


Figura 14: Estructura de una horquilla de DNA (izquierda) y representación de una estructura cruciforme (derecha).

Para el estudio de este tipo de estructuras por RMN, las condiciones óptimas normalmente consisten en trabajar a concentraciones relativamente bajas de muestra, baja fuerza iónica y temperaturas relativamente altas⁴⁶. De este modo, se evitan conformaciones multiméricas que dificultan la determinación estructural de la horquilla^{47,48}. En el capítulo 5 hemos determinado la conformación de una horquilla formada a partir de modificaciones químicas importantes en terapias antisentido.

2.2.3.3. Estructuras cruciformes

Los denominados cruciformes son estructuras en las que las dobles hélices se cruzan, de tal forma que las hebras de cada hélice se intercambian (Figura 14). Estas asociaciones pueden producirse por uniones perfectas, donde todas las bases están apareadas con sus complementarias mediante pares WC, o bien, pueden aparecer *mismatches*.

En el DNA, el principal significado biológico de estas estructuras es la formación de intermedios en procesos de recombinación genética, replicación y resolución de los telómeros⁴⁹⁻⁵². Normalmente son difíciles de estudiar y resolver por RMN porque presentan conformaciones no estándar o con distorsiones. Además, poseen pesos moleculares altos lo que da lugar a mucho solapamiento y mayores anchuras de línea, que hacen muy laboriosa la asignación y el cálculo de

estructura. Por otro lado, estos sistemas tienden a tener una solubilidad bastante baja, lo cual dificulta su estudio en disolución.

2.2.3.4 Hélice triple o tríplex

La hélice triple (o tríplex) es un dúplex al que se le ha unido una tercera hebra en el surco mayor⁵³⁻⁵⁵. El dúplex está generalmente compuesto por una cadena de homopurinas hibridada con una cadena de homopirimidinas. Según la orientación y composición de la tercera cadena existen dos tipos de hélices triples: (I) tríplex paralelo, también conocido como motivo pirimidina (Pyr.Pur*Pyr)^{55,56}, y (II) tríplex antiparalelo o motivo purina (Pyr.Pur*Pur)⁵⁶.

El tríplex paralelo se caracteriza porque las pirimidinas de la tercera cadena se hibridan con las purinas del dúplex diana mediante la formación de enlaces tipo Hoogsteen (Figura 15). Esto da lugar a la formación de triadas C·G⁺C⁺ y T·A⁺T (o U·A⁺U en el caso de RNA). Una de las características fundamentales de este tipo de tríplex, es que las citosinas de la tercera cadena deben estar protonadas en la posición N3 para poder aparearse con el N7 de la guanina. Por ello, su formación es dependiente del pH y está favorecida a pH bajos.

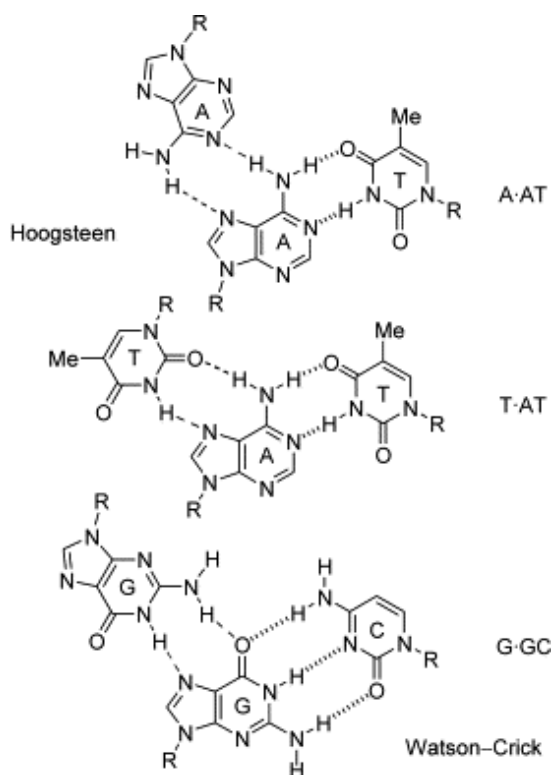


Figura 15: Triadas posibles en triplex paralelos: $T \cdot A^*T$, $C \cdot G^*G$ y $T \cdot A^*A$. La base más a la izquierda está apareada por enlace tipo Hoogsteen, mientras la de la derecha por pares Watson-Crick.

El triplex antiparalelo se forma cuando las purinas de la tercera cadena se hibridan con las purinas del dúplex diana mediante enlaces Hoogsteen reverso (Figura 15). Esta asociación da lugar a triadas $C \cdot G^*G$, $T \cdot A^*A$ o $U \cdot A^*A$ y $T \cdot A^*T$ en el DNA y RNA, lo que permite la inserción de timinas en la tercera cadena compuesta mayoritariamente por purinas. La formación de este tipo de triplex, al contrario que los de tipo paralelo, no depende del pH y normalmente presenta más de una conformación en disolución, dificultando su caracterización mediante RMN.

No está nada claro qué papel juegan estas estructuras en la célula. Se piensa que pueden estar implicadas en la regulación transcripcional o en la inhibición de ciertas enzimas como la “RNA polimerase”, la “DNAase I” y la “RNAase”⁵⁷⁻⁵⁹. La necesidad de hebras ricas en purinas y pirimidinas podría restringir el papel

biológico de estas estructuras a células eucariotas, donde se observa una mayor proporción de este tipo de secuencias⁵⁹.

Otra área que hace muy interesante el estudio de las hélices triples, es la de su uso como agentes terapéuticos. Las aplicaciones de estas estructuras se basan en bloquear la transcripción de ciertos genes de forma específica mediante la hibridación de una tercera cadena a la doble hélice de DNA⁶⁰. Este concepto se denomina estrategia antígeno y se discutirá más extensivamente en el capítulo 3.

2.2.3.5. Hélices cuádruples

Las hélices cuádruples son estructuras formadas por la asociación de cuatro hebras. Existen dos tipos de hélices cuádruples dependiendo de las bases que se asocian para formar los pares de bases: (I) citosinas que dan lugar a los denominados *i-motif* y (II) guaninas que dan lugar a los llamados “*G-quadruplex*”.

Cuádruplex de guaninas (“G-quadruplex”)

Las secuencias ricas en guaninas tienden a asociarse entre sí mediante pares de bases tipo Hoogsteen⁶¹. Cada guanina es aceptor y donador de dos enlaces de hidrógeno (Figura 16). Las secuencias ricas en guaninas, bajo la influencia de cationes monovalentes como Na⁺ o K⁺, se asocian para dar lugar a estructuras compactas en forma de hélices de cuatro cadenas o cuádruplex⁶². El catión se introduce en medio de la tétrada y coordina los oxígenos que se encuentran apuntando al centro de la hélice, favoreciendo y estabilizando este tipo de estructuras.

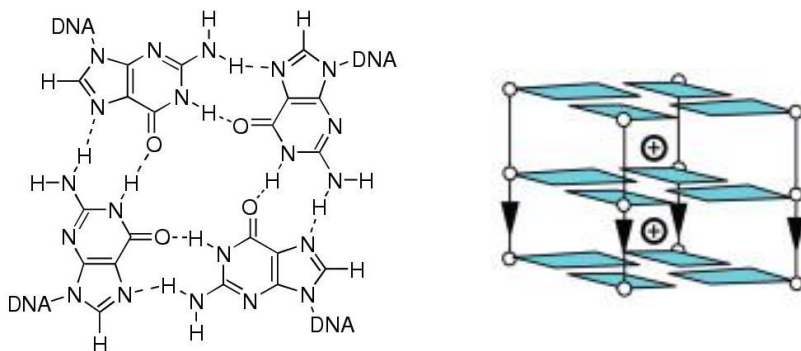


Figura 16: Tetrada de un cuádruplex de guaninas formada por cuatro guaninas unidas por enlaces de hidrógeno tipo Hoogsteen. El catión coordina los oxígenos que apuntan hacia el centro de la tetrada.

La longitud de la secuencia del oligonucleótido determina el tipo de plegamiento del cuádruplex. Cadenas cortas que poseen un único fragmento de tres o cuatro guaninas requieren cuatro cadenas para formar el cuádruplex. Este tipo de cuádruplex se denomina tetramolecular, ya que necesita cuatro hebras para su formación (Figura 17). Además, carece de loops, elementos estructurales necesarios para unir las hebras de guanina entre sí. Secuencias más largas, con dos fragmentos de guaninas continuas separadas por una o más bases de otro tipo, permiten la formación de cuádruplex bimoleculares. Así, dos cadenas polinucleotídicas se asocian para formar un cuádruplex. Finalmente, secuencias que contienen cuatro segmentos de guaninas continuas son capaces de plegarse sobre sí mismas y formar los denominados cuádruplex intramoleculares⁶³⁻⁶⁵.

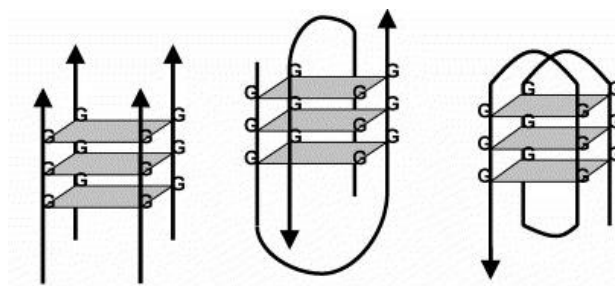


Figura 17: Representación de los cuádruplex tetramoleculares, bimoleculares y intramoleculares respectivamente.

Además de haber distintos tipos de cuádruplex dependiendo del número de hebras que lo forman, existen diferentes tipos de conformaciones dependiendo de la orientación de las hebras. Así, la distribución de las cadenas en el cuádruplex dará lugar a diferencias en los ángulos glicosídicos de las bases (*anti* o *syn*) y en la distribución de los distintos tipos de *loops*. Por ejemplo, si todas las hebras se disponen en la misma dirección 5'-3', el cuádruplex se denomina paralelo. Si por el contrario, en el cuádruplex hay hebras con sentidos opuestos, se denominan antiparalelos. Dentro de los antiparalelos, podemos encontrar los denominados cuádruplex (3+1), que tienen tres cadenas en un sentido y otra en el contrario, y los (2+2) que tienen dos hebras en un sentido y otras dos en el contrario. Además, en los cuádruplex intra- y bimoleculares, la asociación de las cadenas forma distintos tipos de *loops*, que se distribuyen de forma diferente según el tipo de cuádruplex. Mientras el cuádruplex paralelo tan solo tiene *loops* tipo "*propeller*" y las guaninas en *anti*, los cuádruplex antiparalelos tienen todo tipo de *loops* (diagonal, *edge-wise*, ...) (Figura 18) y diferentes combinaciones de guaninas *anti* y *syn*⁶⁵.

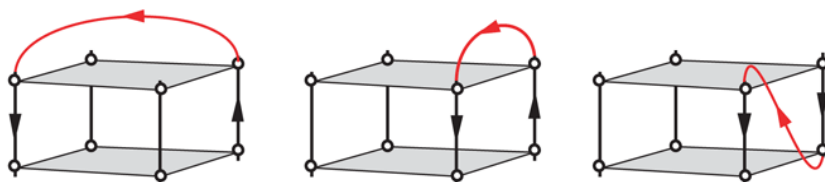


Figura 18: Diferentes tipos de *loops* encontrados en cuádruplex: *diagonal*, *edge-wise* y *propeller* respectivamente.

Finalmente destacar la importancia del catión en la estabilidad y topología de los cuádruplex. El potasio (K^+) es el catión monovalente que más los estabiliza, debido a una mejor coordinación con los oxígenos del interior de la tétrada. Por otro lado, la conformación de cuádruplex también puede cambiar según el catión utilizado, como ocurre en muchas ocasiones con Na^+ y K^+ .

Los cuádruplex, por tanto, son capaces de adoptar conformaciones muy diferentes. Es interesante destacar que esta multitud de topologías tan solo se encuentran en oligonucleótidos de DNA. En RNA, sin embargo, sólo se ha

observado un tipo de cuádruplex paralelo con todas las cadenas en el mismo sentido y con los *loops* “*propeller*”^{66,67}.

Los cuádruplex tienen una gran relevancia biológica. Durante los últimos diez años, se ha observado la presencia de estas estructuras *in vivo*, y la existencia de secuencias ricas en guaninas tanto en el telómero como en secuencias promotoras⁶⁸⁻⁷⁰. Actualmente, se consideran potenciales dianas terapéuticas capaces de inhibir tanto la actividad como la expresión de ciertas proteínas⁷¹. Por ejemplo, la estabilización de los cuádruplex de secuencias teloméricas, mediante la adición de ligandos, es capaz de inhibir la telomerasa, proteína que está presente en el 80% de las células cancerosas y que es fundamental en la elongación de los telómeros^{72,73}. Por otro lado, la formación de cuádruplex en regiones promotoras de oncogenes afecta la expresión de proteínas fundamentales para las células cancerosas. Por ello, el desarrollo de nuevos ligandos y modificaciones químicas, que estabilicen este tipo de estructuras, es de gran importancia para el desarrollo de nuevos fármacos.

I-motif

Son estructuras tetraméricas que se forman por asociación de secuencias ricas en citosinas a pH ácido (pH~5). Inicialmente se descubrió la formación de dúplex paralelos asociados mediante pares $C \cdot C^{+74}$. Posteriormente, estudios por RMN, identificaron la existencia de estructuras de orden superior en las que dos dúplex paralelos se intercalaban entre sí de forma antiparalela (Figura 19)⁷⁵. Los *i-motif*, por tanto, tienen como unidad básica el par $C \cdot C^{+}$, con una citosina protonada en N3 que permite la formación de tres enlaces de hidrógeno (Figura 19). El hecho de que los dúplex estén intercalados permite el apilamiento de los pares $C \cdot C^{+}$, que estabilizan la estructura. Los surcos son muy diferentes a los representados en dúplex, mostrando dos surcos mayores muy anchos y poco profundos y dos surcos menores muy estrechos. Los surcos mayores tienen una anchura aproximada de 16 Å, mientras los surcos menores son de 6 Å^{76,77}. Al igual que en el caso de los cuádruplex, los *i-motif* pueden clasificarse en distintos grupos según el número de hebras que los formen. Así, los tetramoleculares están formados por

cuatro cadenas independientes, los bimoleculares por dos cadenas y los intramoleculares por una sola cadena que se repliega sobre sí misma.

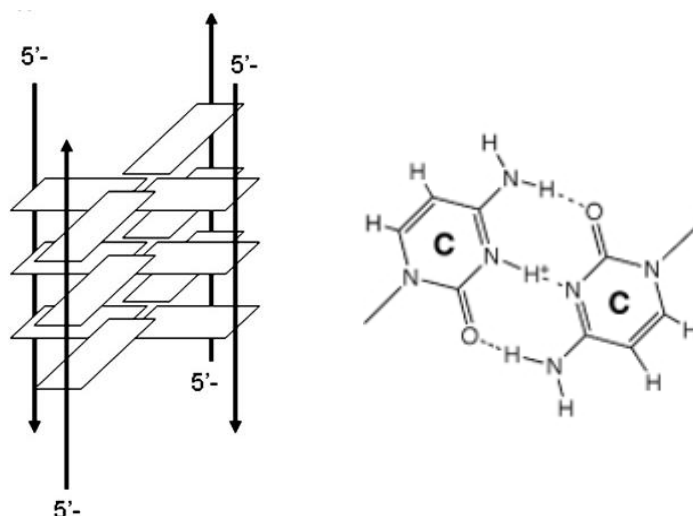


Figura 19: Representación esquemática de la estructura del *i-motif* y del tipo de pares de bases que los forman.

Este tipo de estructuras se ha observado tanto en DNA como en RNA, aunque en RNA son mucho menos estables⁷⁸. Se suelen formar en secuencias ricas en citosinas, normalmente complementarias a las ricas en guaninas que tienden a formar cuádruplex de guaninas. Recientemente, además, se ha observado la formación de *i-motif* en secuencias con muy pocas citosinas⁷⁹. Al contrario que en el caso de los cuádruplex, todavía no existe evidencia experimental que apoye la implicación de este tipo de estructuras en procesos biológicos. *A priori*, uno de los principales problemas es el pH ácido requerido para su formación. El pK_a del N3 de la citosina es 4.3 en el nucleótido, sin embargo, en *i-motif* intramoleculares el pK_a puede ser de hasta 7.5, lo que apoyaría la formación de este tipo de estructuras a pH neutro. Estos resultados sugieren que, al menos en el caso de *i-motif* intramoleculares, el pH ácido no es un problema.

A nivel biológico, la estructura tipo *i-motif* podría ser una posible diana terapéutica teniendo en cuenta la capacidad de las cadenas de policitosinas modificadas para inhibir la proteína “*HIV-1 integrase*” y su localización en secuencias promotoras⁸⁰.

Además, se ha planteado su posible implicación en la función del centrómero y como estructuras transitorias en la elongación del telómero⁸¹.

2.3 Estabilización de las estructuras secundarias de los ácidos nucleicos

La estabilidad de los ácidos nucleicos es un fenómeno complejo, donde juega un papel fundamental la compensación entre factores entálpicos y entrópicos. La hibridación de las cadenas de DNA y RNA es, en general, entrópicamente desfavorable ($\Delta S < 0$ y $-T\Delta S > 0$), efecto que normalmente se compensa con interacciones entálpicamente favorables ($\Delta H < 0$)⁸².

El término ΔS depende de varios factores, como son la flexibilidad de los ángulos glicosídicos y de torsión de las cadenas polinucleotídicas, o el grado de orden/desorden de las moléculas del solvente. En general, las cadenas sencillas en disolución (S_i) tendrán una entropía mayor que las estructuras hibridadas (S_f). El esqueleto y las bases de los dúplex se encuentran mucho más rígidos, y por tanto más ordenados, que las cadenas polinucleotídicas sin hibridar^{44,83}. Este efecto es muy desfavorable entrópicamente con valores de $\Delta S < 0$ y $-T\Delta S > 0$. Los valores tan favorables de entalpía compensan los valores desfavorables de entropía y se obtienen valores de ΔG negativos que representan un proceso espontáneo. Normalmente los valores de energía libre son pequeños, lo que además representa equilibrios fácilmente reversibles.

El término ΔH es muy favorable en los procesos de formación de dúplex, tríplex y demás estructuras secundarias. Las interacciones de apareamiento y apilamiento entre las bases, las interacciones con el medio y otras interacciones electrostáticas⁸⁴ permiten unos valores de $\Delta H < 0$.

La interacción entre las bases, como hemos descrito anteriormente, se debe a la formación de enlaces de hidrógeno entre ellas. La formación de dos o tres enlaces por par de bases favorece mucho la hibridación entre las hebras.

Otro modo de interacción entre bases es la interacción de apilamiento o *stacking*. En disolución acuosa, las bases tienden a apilarse de forma paralela una sobre otra a una distancia aproximada de 3.4 Å⁴⁴. Es un proceso local que viene dado principalmente por las interacciones de una base con sus dos bases vecinas. Estas interacciones están relacionadas con el solapamiento de orbitales π de los anillos aromáticos. Este tipo de interacción es un fenómeno complejo en el que están implicados: (a) la interacción electrostática que origina la disposición de bases apiladas. Ésta maximiza la interacción favorable y minimiza la interacción desfavorable entre las cargas de las bases en contacto, (b) las fuerzas de van der Waals, responsables de la especificidad en el apilamiento, y (c) en medios acuosos, la consideración de efectos hidrófobos, que van a contribuir a la estabilización de la estructura secundaria y terciaria. También en medios acuosos, la doble hélice de DNA favorece interacciones de apilamiento que estabilizan más que las de formación de enlaces de hidrógeno⁸⁵.

El agua y los contraiones son muy importantes a la hora de determinar la estructura de los ácidos nucleicos⁸⁶. La carga negativa de los fosfatos provoca una repulsión electrostática entre ellos que disminuirá por la alta constante dieléctrica del agua y por los contraiones hidratados presentes en la disolución. El grado de hidratación del DNA y RNA no solo es fundamental en la conformación que adquiere, sino además en su estabilidad. Las moléculas de agua forman puentes entre fosfatos sucesivos de una cadena o se encuentran en los grupos PO^{2-} interaccionando preferentemente con los dos oxígenos a la vez. La hidratación en el DNA tipo B y el DNA tipo A es considerablemente distinta, lo que provocará diferentes valores de ΔH . Además, la hidratación y la naturaleza de los contraiones pueden favorecer la formación de estructuras no canónicas como hélices cuádruples o hélices triples⁶².

Por último, mencionar que existen otras interacciones electrostáticas que estabilizan la estructura secundaria y terciaria de los ácidos nucleicos^{87,88}. Se ha demostrado cómo enlaces de hidrógeno entre el grupo hidroxilo del azúcar y el esqueleto estabilizan estructuras tipo A⁸⁹. Además, otras interacciones, como los

enlaces de hidrógeno no convencionales tipo C-H...O, favorecen interacciones entre ácidos nucleicos, ligandos y proteínas^{88,90}.

BIBLIOGRAFÍA

- (1) Blackburn, G. M. G., M.J. *Nucleic Acids in Chemistry and Biology*, IRL Press, Oxford, (1990).
- (2) Altona, C.; Sundaralingam, M. *Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation* **J. Am. Chem. Soc.** (1972), 94, 8205.
- (3) Altona, C.; Sundaralingam, M. *Conformational analysis of the sugar ring in nucleosides and nucleotides. Improved method for the interpretation of proton magnetic resonance coupling constants* **J. Am. Chem. Soc.** (1973), 95, 2333.
- (4) Saenger, W. *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, (1984).
- (5) Donohue, J.; Trueblood, K. N. *Base pairing in DNA* **J. Mol. Biol.** (1960), 2, 363.
- (6) Haschemeyer, A. E. V.; Rich, A. *Nucleoside conformations: an analysis of steric barriers to rotation about the glycosidic bond* **J. Mol. Biol.** (1967), 27, 369.
- (7) Sundaralingam, M. *Structure and conformation of nucleosides and nucleotides and their analogs as determined by X-Ray diffraction* **Ann. N. Y. Acad. Sci.** (1975), 255, 3.
- (8) Wolfenden, R. V. *Tautomeric equilibria in inosine and adenosine* **J. Mol. Biol.** (1969), 40, 307.
- (9) Fasman, G. D. *Handbook of Biochemistry and Molecular Biology*, Cleveland, (1975); Vol. 1.
- (10) Birnbaum, G. I.; Giziewicz, J.; Huber, C. P.; Shugar, D. *Intramolecular hydrogen bonding and acidities of nucleoside sugar hydroxyls. Crystal structure and conformation of O2,2'-anhydro-1- α -D-xylofuranosyluracil* **J. Am. Chem. Soc.** (1976), 98, 4640.
- (11) Pauling, L. *The Nature of the Chemical Bond* Cornell University Press, Ithaca, New York, (1960).
- (12) Kollman, P. *Hydrogen Bonding and Donor—Acceptor Interactions*; Springer US, Springer US, (1977); Vol. 4.
- (13) Kollman, P. A.; Allen, L. C. *Theory of the hydrogen bond* **Chem. Rev.** (1972), 72, 283.
- (14) Crick, F. Watson, J. *Molecular Structure of Nucleic Acids* **Nature** (1953).
- (15) Crick, F.; Watson, J. D. *The Complementary Structure of Deoxyribonucleic Acid* **Proc. R. Soc. Lond. A. Math. Phys.** (1954), 223, 80.
- (16) Rosenberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *RNA double-helical fragments at atomic resolution: II. The crystal structure of sodium guanylyl-3',5'-cytidine nonahydrate* **J. Mol. Biol.** (1976), 104, 145.
- (17) Hoogsteen, K. *The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine* **Acta Crystallogr.** (1963), 16, 907.
- (18) Colominas, C.; Luque, F. J.; Orozco, M. *Tautomerism and Protonation of Guanine and Cytosine. Implications in the Formation of Hydrogen-Bonded Complexes* **J. Am. Chem. Soc.** (1996), 118, 6811.
- (19) Isaksson, J.; Zamaratski, E.; Maltseva, T. V.; Agback, P.; Kumar, A.; Chattopadhyaya, J. *The First Example of a Hoogsteen Basepaired DNA Duplex in Dynamic Equilibrium with a Watson-Crick Basepaired Duplex—A Structural (NMR), Kinetic and Thermodynamic Study* **J. Biomol. Struct. Dyn.** (2001), 18, 783.
- (20) Hakoshima, T.; Fukui, T.; Ikehara, M.; Tomita, K.-i. *Molecular structure of a double helix that has non-Watson—Crick type base pairing formed by 2-substituted poly(A) and poly(U)* **Proc. Natl. Acad. Sci.** (1981), 78, 7309.
- (21) Pasternack, L. B.; Lin, S.-B.; Chin, T.-M.; Lin, W.-C.; Huang, D.-H.; Kan, L.-S. *Proton NMR Studies of 5'-d-(TC)3 (CT)3 (AG)3-3'—A Paperclip Triplex: The Structural Relevance of Turns* **Biophys. J.** (2002), 82, 3170.
- (22) Blommers, M. J. J.; Van De Ven, F. J. M.; Van Der Marel, G. A.; Van Boom, J. H.; Hilbers, C. W. *The three-dimensional structure of a DNA hairpin in solution* **Eur. J. Biochem.** (1991), 201, 33.
- (23) Patikoglou, G. A.; Kim, J. L.; Sun, L.; Yang, S.-H.; Kodadek, T.; Burley, S. K. *TATA element recognition by the TATA box-binding protein has been conserved throughout evolution* **Genes Dev.** (1999), 13, 3217.
- (24) Auffinger, P.; Westhof, E. *Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes* **J. Mol. Biol.** (1999), 292, 467.
- (25) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. *Transient Hoogsteen base pairs in canonical duplex DNA* **Nature** (2011), 470, 498.

- (26) Raghunathan, G.; Miles, H. T.; Sasisekharan, V. *Parallel nucleic acid helices with Hoogsteen base pairing: Symmetry and structure* **Biopolymers** (1994), 34, 1573.
- (27) Donohue, J. *Hydrogen-bonded helical configurations of polynucleotides* **Proc. Natl. Acad. Sci. U. S. A.** (1956), 42, 60.
- (28) Parvathy, V. R.; Bhaumik, S. R.; Chary, K. V. R.; Govil, G.; Liu, K.; Howard, F. B.; Miles, H. T. *NMR structure of a parallel-stranded DNA duplex at atomic resolution* **Nucleic Acids Res.** (2002), 30, 1500.
- (29) Frank-Kamenetskii, M. D.; Mirkin, S. M. *Triplex DNA structures* **Annu. Rev. Biochem.** (1995), 64, 65.
- (30) Crick, F. H. C. *Codon—anticodon pairing: The wobble hypothesis* **J. Mol. Biol.** (1966), 19, 548.
- (31) Sakore, T. D.; Sobell, H. M. *Crystal and molecular structure of a hydrogen-bonded complex containing adenine and hypoxanthine derivatives: 9-Ethyl-8-bromoadenine-9-ethyl-8-bromohypoxanthine* **J. Mol. Biol.** (1969), 43, 77.
- (32) Freese, E. *The difference between Spontaneous and Base-Analogue Induced Mutations of Phage T4* **Proc. Natl. Acad. Sci. U. S. A.** (1959), 45, 622.
- (33) Neidle, S. *Nucleic Acid Structure*; Oxford, (1999).
- (34) Geierstanger, B. H.; Wemmer, D. E. *Complexes of the minor groove of DNA* **Annu. Rev. Biophys. Biomol. Struct.** (1995), 24, 463.
- (35) Manzanares, I.; Cuevas, C.; García-Nieto, R.; Marco, E.; Gago, F. *Advances in the chemistry and pharmacology of ecteinascidins, a promising new class of anticancer agents* **Curr. Med. Chem.: Anti-Cancer Agents** (2001), 1, 257.
- (36) Kwon, Y.; Xi, Z.; Kappen, L. S.; Goldberg, I. H.; Gao, X. *New Complex of Post-Activated Neocarzinostatin Chromophore with DNA: Bulge DNA Binding from the Minor Groove* **Biochemistry** (2003), 42, 1186.
- (37) Richard, W.; Horace, D.; Tsunehiro, T.; Chris, B.; Shoji, T.; Keiichi, I.; Richard, E. D. *Crystal structure analysis of a complete turn of B-DNA* **Nature** (1980), 287, 755.
- (38) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Structure of a B-DNA dodecamer: conformation and dynamics* **Proc. Natl. Acad. Sci.** (1981), 78, 2179.
- (39) Tereshko, V.; Minasov, G.; Egli, M. *The Dickerson-Drew B-DNA Dodecamer Revisited at Atomic Resolution* **J. Am. Chem. Soc.** (1998), 121, 470.
- (40) Tjandra, N.; Tate, S.-i.; Ono, A.; Kainosho, M.; Bax, A. *The NMR Structure of a DNA Dodecamer in an Aqueous Dilute Liquid Crystalline Phase* **J. Am. Chem. Soc.** (2000), 122, 6190.
- (41) Maizels, N.; Gray, L. T. *The G4 Genome* **PLoS Genet** (2013), 9, e1003468.
- (42) Wang, A.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; Van Boom, J. H.; van der Marel, G.; Rich, A. *Molecular structure of a left-handed double helical DNA fragment at atomic resolution* **Nature** (1979), 282, 680.
- (43) Hall, K.; Cruz, P.; Tinoco, I.; Jovin, T. M.; van de Sande, J. H. *'Z-RNA'—a left-handed RNA double helix* **Nature** (1984), 311, 584.
- (44) Kool, E. T. *Preorganization of DNA: Design principles for improving nucleic acid recognition by synthetic oligonucleotides* **Chem. Rev.** (1997), 97, 1473.
- (45) Hald, M.; Pedersen, J. B.; Stein, P. C.; Kirpekar, F.; Jacobsen, J. P. *A comparison of the hairpin stability of the palindromic d (CGCG (A/T) 4CGCG) oligonucleotides* **Nucleic Acids Res.** (1995), 23, 4576.
- (46) Marky, L. A.; Blumenfeld, K. S.; Kozłowski, S.; Breslauer, K. J. *Salt-dependent conformational transitions in the self-complementary deoxydodecanucleotide d (CGCAATTCGCG): Evidence for hairpin formation* **Biopolymers** (1983), 22, 1247.
- (47) Zhou, N.; Germann, M. W.; van de Sande, J. H.; Pattabiraman, N.; Vogel, H. J. *Solution structure of the parallel-stranded hairpin d (T8 <> C4A8) as determined by two-dimensional NMR* **Biochemistry** (1993), 32, 646.
- (48) Germann, M. W.; Vogel, H. J.; Pon, R. T.; van de Sande, J. H. *Characterization of a parallel stranded DNA hairpin* **Biochemistry** (1989), 28, 6220.
- (49) Holliday, R. *A mechanism for gene conversion in fungi* **Genet. Res.** (1964), 5, 282.
- (50) Szostak, J. W.; Blackburn, E. H. *Cloning yeast telomeres on linear plasmid vectors* **Cell** (1982), 29, 245.
- (51) Potter, H.; Dressler, D. *On the mechanism of genetic recombination: electron microscopic observation of recombination intermediates* **Proc. Natl. Acad. Sci.** (1976), 73, 3000.
- (52) Potter, H.; Dressler, D. *In vitro system from Escherichia coli that catalyzes generalized genetic recombination* **Proc. Natl. Acad. Sci.** (1978), 75, 3698.
- (53) Radhakrishnan, I.; Patel, D. J. *DNA triplexes: solution structures, hydration sites, energetics, interactions, and function* **Biochemistry** (1994), 33, 11405.
- (54) Thuong, N. T.; Hélène, C. *Sequence-specific recognition and modification of double-helical DNA by oligonucleotides* **Angew. Chem. Int. Ed.** (1993), 32, 666.
- (55) Felsenfeld, G.; Davies, D. R.; Rich, A. *Formation of a three-stranded polynucleotide molecule* **J. Am. Chem. Soc.** (1957), 79, 2023.

- (56) Chandler, S. P.; Fox, K. R. *Specificity of Antiparallel DNA Triple Helix Formation* **Biochemistry** (1996), 35, 15038.
- (57) Morgan, A.; Wells, R. *Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences* **J. Mol. Biol.** (1968), 37, 63.
- (58) Murray, N. L.; Morgan, A. R. *Enzymatic and Physical Studies on the Triplex dTn· dAn· rUn* **Can. J. Biochem.** (1973), 51, 436.
- (59) Behe, M. J. *An overabundance of long oligopurine tracts occurs in the genome of simple and complex eukaryotes* **Nucleic Acids Res.** (1995), 23, 689.
- (60) Duval-Valentin, G.; Thuong, N. T.; Helene, C. *Specific inhibition of transcription by triple helix-forming oligonucleotides* **Proc. Natl. Acad. Sci.** (1992), 89, 504.
- (61) Sen, D.; Gilbert, W. *Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis* **Nature** (1988).
- (62) Pinnavaia, T.; Marshall, C. L.; Mettler, C. M.; Fisk, C. L.; Miles, H. T.; Becker, E. D. *Alkali metal ion specificity in the solution ordering of a nucleotide, 5'-guanosine monophosphate* **J. Am. Chem. Soc.** (1978), 100, 3625.
- (63) Simonsson, T. *G-quadruplex DNA structures variations on a theme* **Biol. Chem.** (2001), 382, 621.
- (64) Patel, D.; Bouaziz, S.; Kettani, A.; Wang, Y. *Structures of guanine-rich and cytosine-rich quadruplexes formed in vitro by telomeric, centromeric, and triplet repeat disease DNA sequences*, Oxford University Press, Oxford, (1999).
- (65) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Quadruplex DNA: sequence, topology and structure* **Nucleic Acids Res.** (2006), 34, 5402.
- (66) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Crystal structure of parallel quadruplexes from human telomeric DNA* **Nature** (2002), 417, 876.
- (67) Martadinata, H.; Phan, A. T. *Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K⁺ solution* **J. Am. Chem. Soc.** (2009), 131, 2570.
- (68) Sandell, L. L.; Zakian, V. A. *Loss of a yeast telomere: Arrest, recovery, and chromosome loss* **Cell** (1993), 75, 729.
- (69) De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J.-F.; Mergny, J.-L. *Targeting telomeres and telomerase* **Biochimie** (2008), 90, 131.
- (70) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. *Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo* **Nat Struct Mol Biol** (2005), 12, 847.
- (71) Arthanari, H.; Bolton, P. H. *Functional and dysfunctional roles of quadruplex DNA in cells* **Chem. Biol.** (2001), 8, 221.
- (72) Neidle, S. *The structures of quadruplex nucleic acids and their drug complexes* **Curr. Opin. Struct. Biol.** (2009), 19, 239.
- (73) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. *Inhibition of telomerase by G-quartet DNA structures* **Nature** (1991), 350, 718.
- (74) Langridge, R.; Rich, A. *Molecular structure of helical polycytidylic acid* **Nature** (1963), 198, 725.
- (75) Gehring, K.; Leroy, J.-L.; Guéron, M. *A tetrameric DNA structure with protonated cytosine-cytosine base pairs* **Nature** (1993).
- (76) Berger, I.; Kang, C.; Fredian, A.; Ratliff, R.; Moyzis, R.; Rich, A. *Extension of the four-stranded intercalated cytosine motif by adenine•adenine base pairing in the crystal structure of d (CCCAAT)* **Nat. Struct. Mol. Biol.** (1995), 2, 416.
- (77) Cai, L.; Chen, L.; Raghavan, S.; Rich, A.; Ratliff, R.; Moyzis, R. *Intercalated cytosine motif and novel adenine clusters in the crystal structure of the Tetrahymena telomere* **Nucleic Acids Res.** (1998), 26, 4696.
- (78) Snoussi, K.; Nonin-Lecomte, S.; Leroy, J.-L. *The RNA i-motif* **J. Mol. Biol.** (2001), 309, 139.
- (79) Escaja, N.; Viladoms, J.; Garavís, M.; Villasante, A.; Pedrosa, E.; González, C. *A minimal i-motif stabilized by minor groove G: T: G: T tetrads* **Nucleic Acids Res.** (2012), 40, 11737.
- (80) Guo, K.; Pourpak, A.; Beetz-Rogers, K.; Gokhale, V.; Sun, D.; Hurley, L. H. *Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene* **J. Am. Chem. Soc.** (2007), 129, 10220.
- (81) Yong, C.; Konggang, Q.; Chuanqi, Z.; Li, W.; Jinsong, R.; Jiasi, W.; Xiaogang, Q. *Insights into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres* **Nat. Commun.** (2012), 3, 1074.
- (82) Searle, M. S.; Williams, D. H. *On the stability of nucleic acid structures in solution: enthalpy - entropy compensations, internal rotations and reversibility* **Nucleic Acids Res.** (1993), 21, 2051.

- (83) Freier, S. M.; Sugimoto, N.; Sinclair, A.; Alkema, D.; Neilson, T.; Kierzek, R.; Caruthers, M. H.; Turner, D. H. *Stability of XGCGCp, GCGCp, and XGCGCp helices: an empirical estimate of the energetics of hydrogen bonds in nucleic acids* **Biochemistry** (1986), 25, 3214.
- (84) Kool, E. T. *Hydrogen Bonding, Base Stacking, And steric effects in DNA replication* **Annu. Rev. Biophys. Biomol. Struct.** (2001), 30, 1.
- (85) Nelson, J. C.; Saven, J. G.; Moore, J. S.; Wolynes, P. G. *Solvophobically Driven Folding of Nonbiological Oligomers* **Science** (1997), 277, 1793.
- (86) Edelhoch, H.; Osborne Jr, J. C. *The Thermodynamic Basis of the Stability of Proteins, Nucleic Acids, and Membranes*; Academic Press, Academic Press, (1976).
- (87) Berger, I.; Egli, M.; Rich, A. *Inter-strand C-H...O hydrogen bonds stabilizing four-stranded intercalated molecules: stereoelectronic effects of O4' in cytosine-rich DNA* **Proc. Natl. Acad. Sci.** (1996), 93, 12116.
- (88) Horowitz, S.; Trievel, R. C. *Carbon-Oxygen Hydrogen Bonding in Biological Structure and Function* **J. Biol. Chem.** (2012), 287, 41576.
- (89) Auffinger, P.; Westhof, E. *Rules governing the orientation of the 2'-hydroxyl group in RNA* **J. Mol. Biol.** (1997), 274, 54.
- (90) Jiang, L.; Lai, L. *CH...O Hydrogen Bonds at Protein-Protein Interfaces* **J. Biol. Chem.** (2002), 277, 37732.

Capítulo 3

Nucleótidos modificados químicamente

3.1 Terapias antisentido y de RNA interferente: Conceptos básicos y mecanismos

Un elemento fundamental de la medicina occidental es el uso de compuestos químicos como fármacos. Éstos son capaces de interactuar con las dianas terapéuticas, normalmente biomoléculas que producen la enfermedad o están involucradas en ella. Algunas de las terapias más prometedoras actualmente se basan en la actuación del fármaco sobre los ácidos nucleicos¹⁻⁴. El material genético contiene la información necesaria para que se desarrollen los procesos celulares y, por tanto, es una diana terapéutica muy interesante.

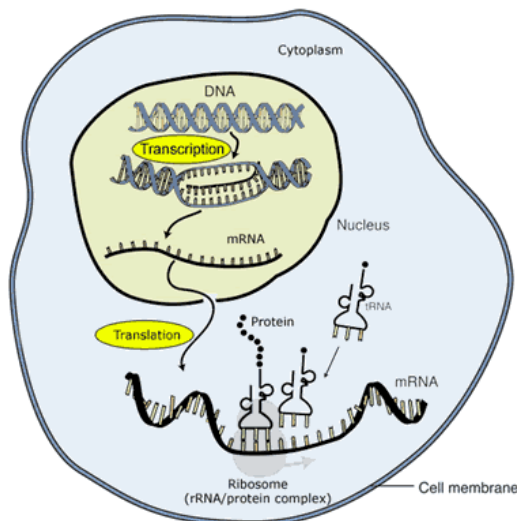


Figura 1: Esquema del proceso de expresión de las proteínas en la célula.

Aunque existen muchos tipos de terapias basadas en la utilización de fármacos que interactúan con los ácidos nucleicos, en esta tesis nos vamos a centrar en algunas de las denominadas terapias de silenciamiento de la expresión génica. El mecanismo de las que aquí se describen se basa en localizar los genes relacionados con la enfermedad, utilizarlos como dianas terapéuticas, e hibridarlos con ácidos nucleicos complementarios⁵⁻⁷. Estas terapias permiten aumentar la eficacia y especificidad del fármaco, y se pueden clasificar en tres grandes grupos dependiendo del mecanismo de acción o del tipo de molécula que se une al gen

(Figura 2): (a) terapias antisentido o antigén, basadas en la utilización de oligonucleótidos como fármacos, (b) terapias de RNA interferente (RNAi) basadas en la activación del complejo RISC (*RNA-induced silencing complex*) mediante un dúplex de RNA y (c) terapias basadas en la utilización de ribozimas modificadas.

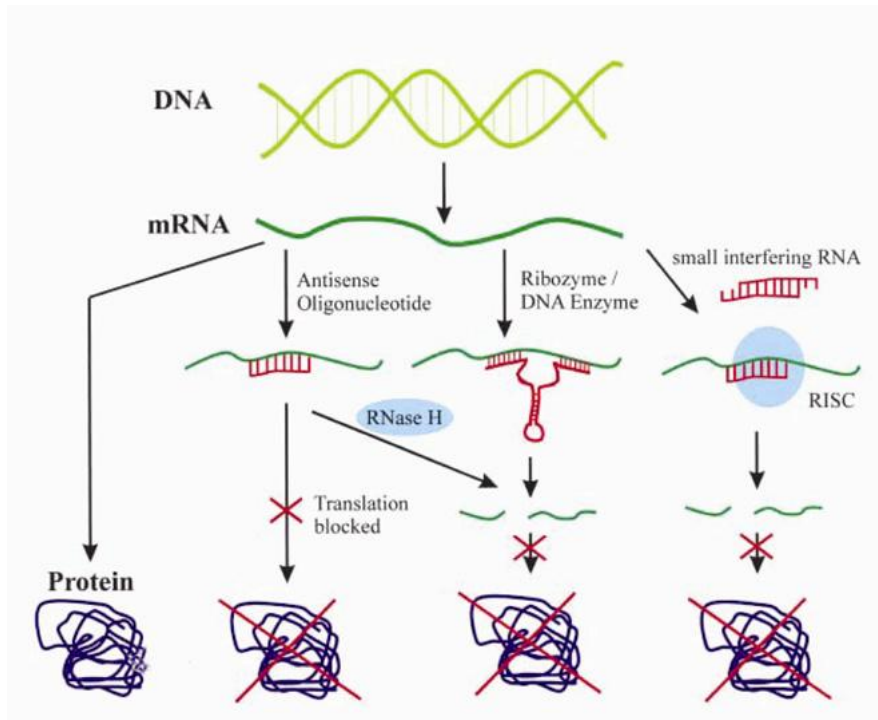


Figura 2: Mecanismo de acción de algunas de las terapias de silenciamiento génico.

3.1.1. Terapias antisentido y antigén

Las terapias antisentido o antigén están basadas en el uso de compuestos oligonucleotídicos, que interactúan con el material genético implicado en la enfermedad. Estos fármacos se denominan oligonucleótidos antisentido o antigén y son pequeños segmentos de DNA o RNA, con longitudes entre 8 y 50 bases, diseñados para unirse específicamente al gen cuya expresión se quiere bloquear. En general, se reserva el término antisentido cuando se hibridan al *mRNA*

mediante pares de bases Watson-Crick, y el término antígeno cuando la diana es el DNA celular y permite la formación de tríplex^{5,8}.

3.1.1.1. Terapias antisentido

El silenciamiento génico se produce a nivel translacional en el citoplasma. El mecanismo de esta terapia se basa en el apareamiento de una cadena oligonucleotídica con el *mRNA* para dar lugar a la formación de un dúplex. El oligonucleótido antisentido tiene una secuencia antiparalela y complementaria a la diana terapéutica (*mRNA*). Teóricamente la hibridación de esta cadena con el *mRNA* provocará la formación de un dúplex que prevendrá el reconocimiento del *mRNA* por parte del complejo ribosomal. Así, la cadena polipeptídica no se expresará, evitándose la formación de la proteína que produce la enfermedad. Aunque teóricamente el proceso parece sencillo, existen multitud de enzimas, además del propio complejo ribosomal, capaces de reparar y separar las cadenas del dúplex.

Aunque existen múltiples mecanismos potenciales por los cuales el oligonucleótido antisentido podría finalizar la actividad de *mRNA*, tan solo dos han sido observados⁷:

Mecanismo de acción de la "RNase H"

Al introducir el oligonucleótido de DNA antisentido complementario al *mRNA*, éstos se hibridan dando lugar a un dúplex DNA:*mRNA*. La *RNase H* (Ribonucleasa H) es ubicua en todo tipo de células y reconoce dúplex híbridos DNA:RNA. Una vez reconocido el dúplex, esta enzima hidroliza la cadena de *mRNA*, liberando la cadena de DNA para que interactúe con otros *mRNA* e inhibiendo, por tanto, la expresión de la proteína asociada a ese gen (Figura 3). Estudios del mecanismo de acción de la *RNase H* determinaron que el tamaño correcto del surco menor, junto con la flexibilidad de la cadena oligonucleotídica y la accesibilidad del 2'-OH en el RNA, eran fundamentales para la actividad de ésta^{9,10}. Así, la *RNase H* solo reconoce dúplex híbridos como es el caso del DNA:RNA. La estructura intermedia de este tipo de dúplex queda reflejada en el surco menor, donde las estructuras de DNA:RNA tienen valores entre 6 y 11 Å. Por el contrario, los homodúplex RNA:RNA y DNA:DNA que forman estructuras canónicas A y B con surcos

menores de 11 y 6 Å respectivamente, y con una flexibilidad conformacional distinta a los dúplex híbridos, no son sustrato de la enzima.

Mecanismo de bloqueo del ribosoma

Mecanismo basado en la prevención de la interacción entre el ribosoma y el *mRNA*. La formación de un dúplex entre el *mRNA* y el oligonucleótido antisentido de DNA o RNA provoca un bloqueo estérico con el ribosoma e inhibe la expresión de la proteína (Figura 3). Uno de los principales problemas de este tipo de terapias es la apertura del dúplex mediante enzimas celulares. Por tanto, un requisito fundamental para evitar este proceso es la alta afinidad entre el oligonucleótido antisentido y el *mRNA*.

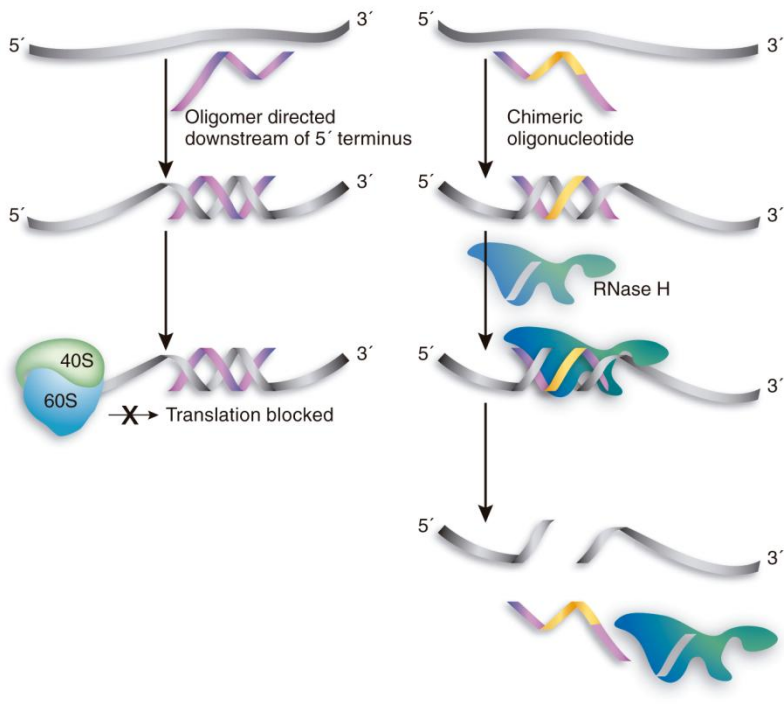


Figura 3: Esquema del mecanismo de bloqueo estérico (izquierda) y de actuación de la RNase H (derecha).

3.1.1.2. Terapias antígeno

El silenciamiento de la expresión génica se produce en el núcleo. La hibridación de una cadena oligonucleotídica a la doble cadena de DNA da lugar a la formación de un tríplex (ver sección 2.2). Este mecanismo, por tanto, permite bloquear la expresión génica a nivel transcripcional sin que se forme el *mRNA*. Los tríplex pueden inhibir la transcripción mediante dos mecanismos. El primero, previniendo la unión de los factores de transcripción a regiones promotoras, y el segundo, evitando la apertura de la doble hélice de DNA. Por tanto, la estabilidad del tríplex es fundamental para el desarrollo de este tipo de terapias. Así, la necesidad de mantener o mejorar la afinidad en la formación del tríplex, la dificultad de encontrar secuencias capaces de formar tríplex y de transportar el oligonucleótido al interior del núcleo, limita mucho la aplicación de las terapias antígeno.

3.1.2 Terapias basadas en la utilización de RNAs interferentes (siRNA)

El descubrimiento de los RNAs interferentes en la célula abrió una nueva puerta para el desarrollo de este tipo de terapias^{11,12}. Uno de los mecanismos de inhibición del gen se basa en activar el complejo multiproteico RISC (*RNA-induced silencing complex*) a través de *siRNAs* (*small interference RNA*). Los *siRNAs* son dúplex de RNA, de 21 nucleótidos por cadena, hibridados mediante pares de bases Watson y Crick y con dos nucleótidos no apareados en cada extremo 3' de las hebras. Éstos se pueden generar en la célula de forma innata, o bien, ser introducidos en el organismo como terapias de silenciamiento de la expresión génica (Figura 4).

En el *siRNA* una de las cadenas se denomina guía (o “*antisense*”) y la complementaria pasajera (o “*sense*”). La cadena guía, a su vez, es complementaria al *mRNA* que se quiere hidrolizar y permitirá su reconocimiento por parte del complejo RISC. Una vez el *siRNA* se une al complejo RISC, éste hidroliza la cadena pasajera. La cadena guía sigue intacta y unida al complejo, lo que permitirá el reconocimiento del *mRNA* por complementariedad de bases. Esta unión al complejo permitirá la hidrólisis del *mRNA*, inhibiéndose así la expresión de la proteína (Figura 4)^{13,14}.

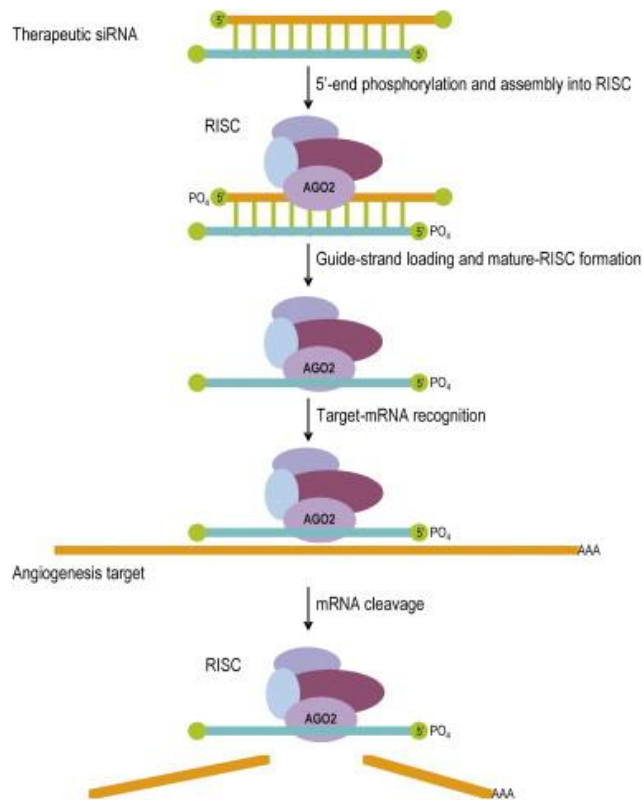


Figura 4: Mecanismo de acción de la terapia basada en los siRNAs.

El mecanismo de acción del complejo RISC todavía no se conoce en su totalidad. Sin embargo, parece que la termodinámica del dúplex^{15,16}, la presencia del grupo fosfato en el extremo 5'¹⁷ y la secuencia en este mismo extremo (los nucleótidos U y A favorecen la interacción con respecto a C y G)¹⁸ son muy importantes a la hora de definir la hebra guía y la pasajera por parte del complejo RISC. Además, el surco mayor del dúplex parece fundamental a la hora de activar este tipo de complejos^{19,20}. Así, en la célula, este mecanismo generalmente se observa en presencia de dúplex de RNA con una estructura canónica tipo A. Los dúplex híbridos de DNA:RNA o DNA:DNA, con estructuras intermedias o tipo B normalmente no activan el complejo RISC.

3.1.3. Terapias basadas en la utilización de ribozimas

Se utiliza el término ribozima para definir enzimas de RNA con actividad catalítica. Éstas son capaces de reconocer cadenas sencillas de RNA y de romper sus enlaces fosfodiéster²¹. El principal problema de estas terapias es que requieren la introducción de moléculas con un peso molecular muy elevado.

3.2. Nucleótidos modificados químicamente

Aunque teóricamente todos estos mecanismos parecen simples, la degradación de las cadenas de DNA y RNA externas al organismo dificulta la aplicación de estas terapias. Las endo- y exonucleasas son las enzimas encargadas de hidrolizar estas cadenas.

Para intentar paliar este problema, se ha desarrollado una gran cantidad de nucleótidos modificados químicamente que mimetizan los oligonucleótidos naturales, y que a su vez, “engañan” a la célula para que los incorpore a su organismo^{7,22,23}. En los últimos 20 años, se ha desarrollado una amplia cantidad de modificaciones que han permitido importantes mejoras en el diseño de oligonucleótidos antisentido y la comercialización del primer fármaco basado en este tipo de terapias. Así, el denominado Vitravene se utiliza para tratar la *AIDS-related CMV retinitis*^{24,25}, mientras otro compuesto antisentido (Kynamro) ha sido aprobado por la FDA. Además, otros 50 compuestos aproximadamente se encuentran actualmente en ensayos clínicos^{2,14}.

Algunos de los requisitos que deben cumplir los oligonucleótidos modificados para terapias antisentido o RNA interferente son: aumentar la resistencia a nucleasas, mantener la selectividad por el RNA, favorecer la hidrólisis del *mRNA*, mantener la incorporación celular y evitar la toxicidad. Para lograr estos objetivos se han desarrollado tres tipos de modificaciones: en los grupos fosfato, en los azúcares (especialmente en la posición 2' de la ribosa) y en las bases (Figura 5).

Las nucleasas actúan sobre el esqueleto (azúcar-fosfato) del nucleótido. Esta propiedad ha hecho que, en general, las modificaciones en estas posiciones

favorezcan oligonucleótidos resistentes a nucleasas. Las bases, sin embargo, son las responsables de la selectividad en el reconocimiento entre cadenas por lo que modificarlas químicamente no es primordial para estas aplicaciones. Además, no suele evitar la degradación por nucleasas²⁶.

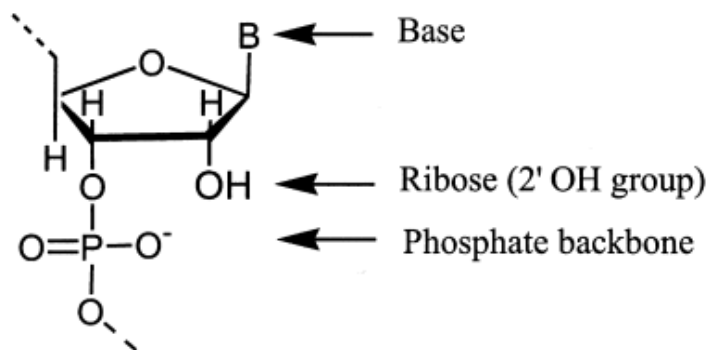


Figura 5: Sitios para las modificaciones químicas en la ribosa. B denota una de las cinco bases nitrogenadas.

3.2.1 Modificaciones de primera generación

La primera generación de modificaciones se encuentra en los grupos fosfato. La modificación más representativa de este grupo es la de los fosforotioatos (PS), en la que uno de los oxígenos del grupo fosfato se substituye por un átomo de azufre (Figura 6)²⁷. El PS-DNA no solo es capaz de aumentar la resistencia a nucleasas, sino que también se une al *mRNA* mediante pares Watson y Crick, es sustrato de la *RNase H* y se introduce en la célula con una serie de propiedades farmacocinéticas muy interesantes²⁸. Entre las desventajas, por otro lado, están la toxicidad y la reducción de afinidad con el *mRNA*²⁹. Tratando de paliar estos problemas se diseñaron una serie de derivados del PS-DNA, como el metil fosfonato, que a su vez provocaron otro tipo de desventajas.

3.2.2 Modificaciones de segunda generación

Muchos de los problemas que planteaban los fosforotioatos (PS) se solventaron en la segunda generación de modificaciones, basadas en la utilización de O-alquilaciones en la posición 2' del azúcar de ribosa (Figura 6). El 2'-O-methyl (OMe) y el 2'-O-methoxy-ethyl RNA (MOE) son las modificaciones más representativas de este grupo. Este tipo de moléculas disminuye la toxicidad y aumenta la afinidad por el *mRNA*^{30,31}.

Estas propiedades, tan deseadas para el desarrollo de fármacos antisentido, se contrarrestaban con el hecho de que los 2'-O-alkyl RNAs no son sustrato de la *RNase H*. Como ya hemos mencionado, estudios del mecanismo de acción de la *RNase H* determinaron que el surco menor, la flexibilidad de la cadena oligonucleotídica, y la accesibilidad del 2'-OH en el RNA son fundamentales para la actividad de ésta. Las 2'-O-alquilaciones, sin embargo, no mantienen estas propiedades, impidiendo su aplicación en terapias antisentido¹⁰.

Aunque este tipo de compuestos no son sustratos de la *RNase H*, el compuesto 2'-O-methoxy-ethyl-RNA es eficaz en procesos de bloqueo estérico³², mientras el compuesto 2'-O-methyl RNA es capaz de activar el complejo RISC y mejorar el efecto de silenciamiento de la expresión génica con respecto al RNA, en ciertas condiciones³³.

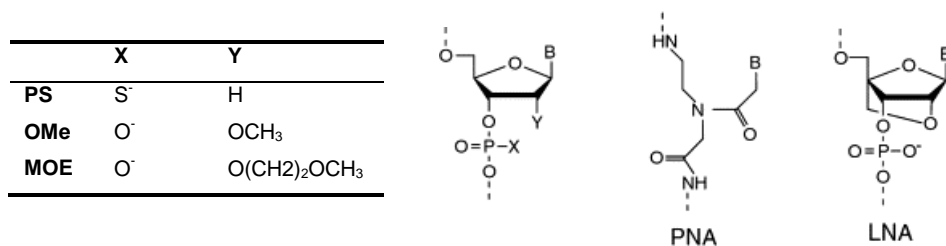


Figure 6: Estructura de algunos de los nucleótidos modificados analizados en este capítulo. PS, OMe, MOE, PNA y LNA.

3.2.3 Modificaciones de tercera generación

En los últimos años, el diseño de nuevos compuestos para mejorar la afinidad al RNA, la resistencia a nucleasas y las propiedades farmacocinéticas, se ha incrementado enormemente. En analogía con los términos “primera generación” para el PS-DNA y “segunda generación” para los 2'-*O-alkyl*-RNAs, estos nuevos nucleótidos se pueden englobar en la llamada “tercera generación”. Se han desarrollado, tanto análogos de DNA y RNA con modificaciones en el azúcar y los grupos fosfato, como nucleótidos con estructuras químicas totalmente diferentes. En esta sección describiremos algunos de los ejemplos más prometedores de esta generación.

Entre estos nucleótidos modificados podemos encontrar los PNAs (*peptide nucleic acids*), donde el esqueleto (azúcar-fosfato) de los nucleótidos se substituye por poliamidas unidas entre sí (Figura 6). Los PNAs tienen una gran afinidad y estabilidad biológica, pero no activan la *RNase H*. Por otro lado, el hecho de que sean moléculas neutras dificulta la solubilidad y la introducción de éstas en la célula. Sin embargo, estos problemas podrían solucionarse uniando las cadenas de PNA a moléculas cargadas negativamente, como lípidos o ciertas proteínas que se introducen fácilmente en la célula^{34,35}.

Otro tipo de modificación es la que da lugar a los LNA (*Loked nucleic acid*). Estos nucleótidos contienen un puente metileno entre el oxígeno 2' del azúcar y el carbono 4' que los bloquea en conformación norte (Figura 6). Al hibridarse con el RNA, esta restricción conformacional favorece la formación de estructuras tipo A. Por tanto, los dúplex LNA:RNA no son sustratos de la *RNase H* e impiden su aplicación en este tipo de terapias antisentido³⁶. Sin embargo, estas propiedades conformacionales permiten activar el complejo RISC^{37,38}. Además, estos nucleótidos aumentan muchísimo la afinidad por el RNA y su resistencia a nucleasas³⁹, lo que los convierte en compuestos muy prometedores como fármacos.

Los azúcares tipo arabinosa (ANA) se caracterizan por ser epímeros del RNA. El grupo hidroxilo en C2' se encuentra en posición contraria (Figura 7), dando lugar a un nucleótido modificado con propiedades totalmente distintas a las del RNA. Este

tipo de nucleótidos forman dúplex híbridos ANA:RNA que son sustratos de la *RNase H*^{40,41}. Sorprendentemente, también se ha observado como, mediante polimerasas diseñadas, es posible generar cadenas de ANA a partir de secuencias de DNA⁴². Aunque la replicación se lleva a cabo con una eficacia mucho menor a la encontrada en DNA, esta propiedad abre nuevas puertas en el entendimiento de los ácidos nucleicos, especialmente en su comportamiento e interacción con proteínas. Además, la capacidad de transmitir la información a partir de nucleótidos de ANA plantea la cuestión evolutiva de por qué nuestro material genético está basado en DNA y RNA y no en otro tipo de nucleótidos.

Existen, además, otros muchos nucleótidos modificados, como el CeNA⁴³ (*Ciclohexene*), tcDNA⁴⁴ (*tríciclo-DNA*), MF⁴⁵ (*Morpholino*), etc. que han sido estudiados en profundidad y también tienen propiedades muy prometedoras.

3.2.4. Nucleótidos fluorados en la posición C2' del azúcar

Especial relevancia tienen los nucleótidos fluorados en la posición C2'. La importancia de los compuestos fluorados a nivel farmacológico ha crecido enormemente en los últimos 20 años, y actualmente engloban el 20% de los medicamentos comerciales^{46,47}. Así, los nucleótidos fluorados 2'-F-ANA (2'-Deoxy-2'-fluoro- β -D-arabinose) y 2'-F-RNA (2'-deoxy-2'-fluoro-RNA) tienen potenciales aplicaciones biológicas, mientras el nucleótido dFdC (2'-deoxy-2'-difluorodeoxycytidine) ya es comercial y se utiliza en tratamientos anticancerígenos.

Los nucleótidos de 2'-F-ANA son análogos fluorados de las arabinosas (ANA) (Figura 7). Éstos poseen algunas características muy similares a los ANA y otras completamente opuestas. Entre las similitudes encontramos que son sustratos de la *RNase H* al unirse al RNA⁴¹, son resistentes a nucleasas⁴⁸ y también se pueden generar cadenas de 2'-F-ANA a partir de DNA y viceversa en presencia de cierto tipo de polimerasas artificiales⁴². Sin embargo, entre las diferencias encontramos que la afinidad de 2'-F-ANA suele ser mucho mayor que la de ANA. Los efectos estructurales que provocan estos nucleótidos en dúplex y cuádruplex, y que

explican esta diferencia de estabilidad, se explicarán en los capítulos de resultados 5, 6 y 9.

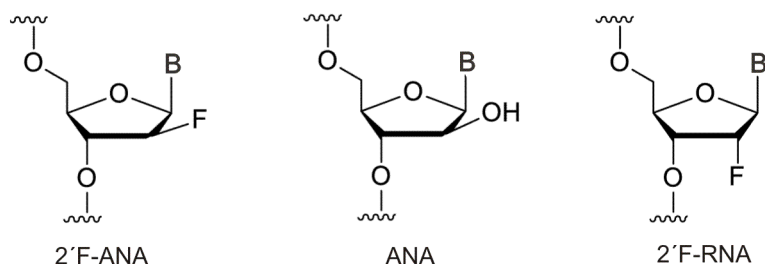


Figura 7: Esquema del nucleótido 2'-F-ANA (izquierda) y ANA (medio) y 2'-F-RNA (derecha).

Por otro lado, los 2'-F-RNAs son epímeros de 2'-F-ANA y análogos fluorados del RNA (Figura 7). Este nucleótido tiende a formar estructuras tipo A, por lo que no es sustrato de la *RNase H*⁴⁹. Sin embargo, estas propiedades conformacionales, junto con el aumento en la afinidad por el RNA y su resistencia a nucleasas⁵⁰, lo convierten en un sustrato perfecto para activar el complejo RISC. El estudio detallado de algunos de los efectos estructurales provocados por esta modificación, y que afectan en gran medida a su afinidad por el RNA, se discuten en el capítulo 7.

Finalmente mencionar el compuesto dFdC (2'-deoxy-2',2'-difluorodeoxycytidine), también denominado Gemcitabina (Figura 8). Este nucleótido posee dos átomos de flúor en la posición C2' y se utiliza como agente quimioterapéutico. El mecanismo de acción de este fármaco, aunque no se comprende en su totalidad, no se basa en ninguna de las terapias de silenciamiento de la expresión génica mencionadas anteriormente. El mecanismo de acción más aceptado actualmente se basa en que el nucleótido dFdC se introduce en la cadena de DNA durante su elongación e impide el alargamiento completo del DNA⁵¹. Esto da lugar a la muerte celular por apoptosis. Aunque el compuesto es comercial y de gran interés biológico, tan sólo existe un estudio estructural de este compuesto en un fragmento de DNA:DNA, donde se observó que la distinta distribución de las cargas en el dFdC podría ser importante a la hora impedir la elongación del DNA por parte de la DNA polimerasa⁵². Para ahondar más en las distorsiones causadas

por el dFdC en otro tipo de sistemas, en el capítulo 8 discutiremos los efectos estructurales de esta modificación en un dúplex híbrido DNA:RNA.

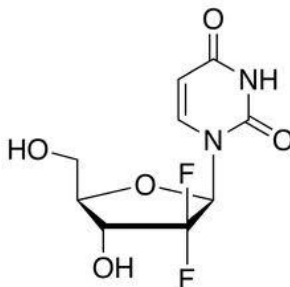


Figura 8: Esquema del nucleótido dFdC.

3.2.5. Modificaciones en las bases

Las bases se encargan del reconocimiento entre cadenas oligonucleotídicas para la formación de estructuras secundarias. Las modificaciones en las bases están diseñadas, en general, para aumentar la afinidad entre las cadenas complementarias. Este efecto se puede favorecer de dos maneras, aumentando el apilamiento entre las bases o aumentando la fuerza del par de bases²⁶. Como ejemplo de modificaciones que aumentan el apilamiento entre bases, se encuentran las modificaciones en la posición C5 de las pirimidinas y las 7-deazapurinas. Las modificaciones que incrementan la fortaleza entre las bases normalmente aumentan el número de enlaces de hidrógeno, por ejemplo, las bases tipo 2-aminoadenina.

Algunas de las propiedades más características de estas modificaciones son: el encontrarse apuntando al surco mayor del dúplex, el provocar, en muchas ocasiones, interacciones catiónicas que aumentan la estabilidad del dúplex y el no afectar la conformación del azúcar. Sin embargo, no suelen conferir resistencia a nucleasas, y aunque en algunas ocasiones este tipo de modificaciones activan el complejo RISC⁵³, en general, no son la estrategia más utilizada en el desarrollo de modificaciones antisentido.

3.3. Estabilidad de los oligonucleótidos modificados

Como hemos podido observar a lo largo de este capítulo, existen múltiples factores que influyen en la posible aplicación de los nucleótidos modificados en terapias de silenciamiento de la expresión génica. Uno de los más importantes es la afinidad de la cadena modificada por el *mRNA*. La estabilización de la estructura secundaria de los ácidos nucleicos depende de una serie de factores entrópicos y entálpicos (ver sección 2.3). Por ello, para que los oligonucleótidos modificados tengan una afinidad apropiada por el *mRNA*, será necesario modular la contribución de cada uno de estos factores a la estabilidad final del sistema.

3.3.1. Factor entrópico: Preorganización

Al hibridarse dos cadenas de DNA o RNA la entropía del estado final suele ser menor que la del estado inicial, dando lugar a valores de ΔS negativos y valores de $-T\Delta S$ positivos. Por tanto, la formación de las estructuras secundarias es, generalmente, entrópicamente desfavorable (sección 2.3).

Uno de los métodos para aumentar la estabilidad de los ácidos nucleicos es hacer menos negativo el término ΔS ^{54,55}. Para ello, se puede disminuir la entropía del estado inicial creando oligonucleótidos modificados más ordenados o preorganizados. Así, mediante modificaciones químicas, moléculas flexibles como el DNA o RNA pueden preorganizarse para optimizar la hibridación. Es importante destacar, que para provocar este efecto, la preorganización de los nucleótidos debe darse en la conformación adecuada para la unión.

3.3.2. Factor entálpico

El factor entálpico, en general, estabiliza las estructuras secundarias de los ácidos nucleicos. Como ya hemos comentado en la sección 2.3 este factor viene dado por un conjunto de interacciones que se favorecen al formarse las estructuras secundarias. Por tanto, para aumentar el efecto de este factor (ΔH) se puede aumentar la fuerza del par de bases, el apilamiento de éstas, modificar la carga global del polinucleótido o favorecer otro tipo de interacciones electrostáticas.

- *Aumentar la fuerza del par de bases*: Normalmente se favorece con modificaciones que aumentan la acidez del dador o la basicidad del aceptor en los enlaces de hidrógeno.

- *Mejorar el apilamiento entre las bases*: Este efecto se consigue aumentando la polarización de las bases, y por tanto, la intensidad de las interacciones π - π (ver sección 2.3).

Es muy importante destacar que mejorar el apilamiento entre las bases no solo influye a la hora de estabilizar la estructura secundaria, sino que además puede provocar un efecto de preorganización en la cadena sencilla.

- *Efectos electrostáticos*: Se puede tratar de evitar la interacción repulsiva entre los grupos fosfato cargados negativamente, aumentar la interacción del esqueleto con las moléculas de agua del medio, o favorecer otro tipo de interacciones electrostáticas que estabilicen la estructura secundaria.

La contribución de cada una de estas interacciones a la estabilidad final de la molécula dependerá principalmente del tipo de estructura secundaria que se forme. Así, la contribución entálpica y entrópica a la estabilidad del sistema variará si la estructura final es una estructura de doble hélice (A, B o Z), o una estructura no canónica, como son las hélices triples o hélices cuádruples (*i-motif* o cuádruplex de guaninas).

BIBLIOGRAFÍA

- (1) Chaires, J. B. *Drug—DNA interactions* **Curr. Opin. Struct. Biol.** (1998), 8, 314.
- (2) Krugh, T. R. *Drug-DNA interactions* **Curr. Opin. Struct. Biol.** (1994), 4, 351.
- (3) Gallego, J.; Varani, G. *Targeting RNA with small-molecule drugs: therapeutic promise and chemical challenges* **Accounts Chem. Res.** (2001), 34, 836.
- (4) Hermann, T.; Westhof, E. *RNA as a drug target: chemical, modelling, and evolutionary tools* **Curr. Opin. Biotechnol.** (1998), 9, 66.
- (5) Aboul-Fadl, T. *Antisense oligonucleotides: the state of the art* **Curr. Med. Chem.** (2005), 12, 2193.
- (6) Bennett, C. F.; Swayze, E. E. *RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform* **Annu. Rev. Pharmacol. Toxicol.** (2010), 50, 259.
- (7) Kurreck, J. *Antisense technologies* **Eur. J. Biochem.** (2003), 270, 1628.
- (8) Uhlmann, E.; Peyman, A. *Antisense oligonucleotides: a new therapeutic principle* **Chem. Rev.** (1990), 90, 543.
- (9) Nakamura, H.; Oda, Y.; Iwai, S.; Inoue, H.; Ohtsuka, E.; Kanaya, S.; Kimura, S.; Katsuda, C.; Katayanagi, K.; Morikawa, K. *How does RNase H recognize a DNA-RNA hybrid?* **Proc. Natl. Acad. Sci. U.S.A.** (1991), 88, 11535.

- (10) Zamaratski, E.; Pradeepkumar, P. I.; Chattopadhyaya, J. *A critical survey of the structure-function of the antisense oligo/RNA heteroduplex as substrate for RNase H* **J. Biochem. Bioph. Methods** (2001), 48, 189.
- (11) Hammond, S. M.; Caudy, A. A.; Hannon, G. J. *Post-transcriptional gene silencing by double-stranded RNA* **Nat. Rev. Gen.** (2001), 2, 110.
- (12) McManus, M. T.; Sharp, P. A. *Gene silencing in mammals by small interfering RNAs* **Nat. Rev. Gen.** (2002), 3, 737.
- (13) Zhang, J.; Hua, Z. *Targeted gene silencing by small interfering RNA-based knock-down technology* **Curr. Pharm. Biotechno.** (2004), 5, 1.
- (14) Templeton, N. S. *Gene and cell therapy: therapeutic mechanisms and strategies*; CRC Press, (2004).
- (15) Khvorova, A.; Reynolds, A.; Jayasena, S. D. *Functional siRNAs and miRNAs exhibit strand bias* **Cell** (2003), 115, 209.
- (16) Schwarz, D. S.; Hutvagner, G.; Du, T. T.; Xu, Z. S.; Aronin, N.; Zamore, P. D. *Asymmetry in the assembly of the RNAi enzyme complex* **Cell** (2007), 131, 30.
- (17) Schwarz, D. S.; Hutvagner, G.; Haley, B.; Zamore, P. D. *Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways* **Mol. Cell** (2002), 10, 537.
- (18) Frank, F.; Sonenberg, N.; Nagar, B. *Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2* **Nature** (2010), 465, 818.
- (19) Chiu, Y.-L.; Rana, T. M. *RNAi in human cells: basic structural and functional features of small interfering RNA* **Mol. Cell** (2002), 10, 549.
- (20) Chiu, Y.-L.; Rana, T. M. *siRNA function in RNAi: a chemical modification analysis* **RNA** (2003), 9, 1034.
- (21) Lewin, A. S.; Hauswirth, W. W. *Ribozyme gene therapy: applications for molecular medicine* **Trends Mol. Med.** (2001), 7, 221.
- (22) Deleavey, G. F.; Damha, M. J. *Designing chemically modified oligonucleotides for targeted gene silencing* **Chem. Biol.** (2012), 19, 937.
- (23) Chaput, John C.; Yu, H.; Zhang, S. *The Emerging World of Synthetic Genetics* **Chem. Biol.** (2012), 19, 1360.
- (24) Marwick, C. *First "Antisense" Drug Will Treat CMV Retinitis* **Jama-J. Am. Med. Assoc.** (1998), 280, 871.
- (25) Crooke, S. In *Drug Delivery Systems in Cancer Therapy*; Brown, D., Ed.; Humana Press: 2004, p 311.
- (26) Herdewijn, P. *Heterocyclic modifications of oligonucleotides and antisense technology* **Antisense Nucleic Acid Drug Dev.** (2000), 10, 297.
- (27) Eckstein, F. *Phosphorothioate oligodeoxynucleotides: what is their origin and what is unique about them?* **Antisense Nucleic Acid Drug Dev.** (2000), 10, 117.
- (28) Crooke, S. T. *Progress in antisense technology: the end of the beginning* **Methods Enzymol.** (2000), 313, 3.
- (29) Levin, A. A. *A review of issues in the pharmacokinetics and toxicology of phosphorothioate antisense oligonucleotides* **Biochim. Biophys. Acta-Gen. Struct. Expression** (1999), 1489, 69.
- (30) Crooke, S. T.; Lemonidis, K. M.; Neilson, L.; Griffey, R.; Lesnik, E. A.; Monia, B. P. *Kinetic Characteristics of escherichia-coli RNase H1- Cleavage of various antisense oligonucleotide-RNA Duplexes* **Biochem. J.** (1995), 312, 599.
- (31) Kurreck, J.; Wyszko, E.; Gillen, C.; Erdmann, V. A. *Design of antisense oligonucleotides stabilized by locked nucleic acids* **Nucleic Acids Res.** (2002), 30, 1911.
- (32) Baker, B. F.; Lot, S. S.; Condon, T. P.; ChengFlournoy, S.; Lesnik, E. A.; Sasmor, H. M.; Bennett, C. F. *2'-O-(2-methoxy)ethyl-modified anti-intercellular adhesion molecule 1 (ICAM-1) oligonucleotides selectively increase the ICAM-1 mRNA level and inhibit formation of the ICAM-1 translation initiation complex in human umbilical vein endothelial cells* **J. Biol. Chem.** (1997), 272, 11994.
- (33) Allerson, C. R.; Sioufi, N.; Jarres, R.; Prakash, T. P.; Naik, N.; Berdeja, A.; Wanders, L.; Griffey, R. H.; Swayze, E. E.; Bhat, B. *Fully 2'-modified oligonucleotide duplexes with improved in vitro potency and stability compared to unmodified small interfering RNA* **J. Med. Chem.** (2005), 48, 901.
- (34) Larsen, H. J.; Bentin, T.; Nielsen, P. E. *Antisense properties of peptide nucleic acid* **Biochim. Biophys. Acta-Gen. Struct. Expression** (1999), 1489, 159.
- (35) Elayadi, A. N.; Corey, D. R. *Application of PNA and LNA oligomers to chemotherapy* **Curr. Opin. Invest. Drugs** (2001), 2, 558.
- (36) Bondensgaard, K.; Petersen, M.; Singh, S. K.; Rajwanshi, V. K.; Kumar, R.; Wengel, J.; Jacobsen, J. P. *Structural studies of LNA : RNA duplexes by NMR: Conformations and implications for RNase H activity* **Chem.-Eur. J.** (2000), 6, 2687.
- (37) Deleavey, G. F.; Watts, J. K.; Damha, M. J. *Chemical modification of siRNA* **Curr. Protoc. Nucleic Acid Chem.** (2009), Chapter 16, Unit 16.3.
- (38) Braasch, D. A.; Jensen, S.; Liu, Y. H.; Kaur, K.; Arar, K.; White, M. A.; Corey, D. R. *RNA interference in mammalian cells by chemically-modified RNA* **Biochemistry** (2003), 42, 7967.

- (39) Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen, C. E.; Wengel, J. *LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition* **Tetrahedron** (1998), 54, 3607.
- (40) Noronha, A. M.; Wilds, C. J.; Lok, C.-N.; Viazovkina, K.; Arion, D.; Parniak, M. A.; Damha, M. J. *Synthesis and Biophysical Properties of Arabinonucleic Acids (ANA): Circular Dichroic Spectra, Melting Temperatures, and Ribonuclease H Susceptibility of ANA-RNA Hybrid Duplexes* **Biochemistry** (2000), 39, 7050.
- (41) Damha, M. J.; Wilds, C. J.; Noronha, A.; Brukner, I.; Borkow, G.; Arion, D.; Parniak, M. A. *Hybrids of RNA and arabinonucleic acids (ANA and 2'-F-ANA) are substrates of ribonuclease h* **J. Am. Chem. Soc.** (1998), 120, 12976.
- (42) Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P. *Synthetic Genetic Polymers Capable of Heredity and Evolution* **Science** (2012), 336, 341.
- (43) Wang, J.; Verbeure, B.; Luyten, I.; Lescrinier, E.; Froeyen, M.; Hendrix, C.; Rosemeyer, H.; Seela, F.; Van Aerschot, A.; Herdewijn, P. *Cyclohexene nucleic acids (CeNA): Serum stable oligonucleotides that activate RNase H and increase duplex stability with complementary RNA* **J. Am. Chem. Soc.** (2000), 122, 8595.
- (44) Steffens, R.; Leumann, C. J. *Tricycle-DNA: A phosphodiester-backbone based DNA analog exhibiting strong complementary base-pairing properties* **J. Am. Chem. Soc.** (1997), 119, 11548.
- (45) Heasman, J. *Morpholino oligos: Making sense of antisense?* **Dev. Biol.** (2002), 243, 209.
- (46) Filler, R.; Saha, R. *Fluorine in medicinal chemistry: a century of progress and a 60-year retrospective of selected highlights* **Future Med. Chem.** (2009), 1, 777.
- (47) O'Hagan, D. *Fluorine in health care* **Organofluorine containing blockbuster drugs **J. Fluor. Chem.** (2010), 131, 1071.**
- (48) Kalota, A.; Karabon, L.; Swider, C. R.; Viazovkina, E.; Elzagheid, M.; Damha, M. J.; Gewirtz, A. M. *2'-deoxy-2'-fluoro-beta-D-arabinonucleic acid (2'-F-ANA) modified oligonucleotides (ON) effect highly efficient, and persistent, gene silencing* **Nucleic Acids Res** (2006), 34, 451.
- (49) Ikeda, H.; Fernandez, R.; Barchi, J. J.; Huang, X.; Marquez, V. E.; Wilk, A. *The effect of two antipodal fluorine-induced sugar puckers on the conformation and stability of the Dickerson-Drew dodecamer duplex [d(CGCGAATTCGCG)]₂* **Nucleic Acids Res.** (1998), 26, 2237.
- (50) Deleavey, G. F.; Watts, J. K.; Alain, T.; Robert, F.; Kalota, A.; Aishwarya, V.; Pelletier, J.; Gewirtz, A. M.; Sonenberg, N.; Damha, M. J. *Synergistic effects between analogs of DNA and RNA improve the potency of siRNA-mediated gene silencing* **Nucleic Acids Res.** (2010), 38, 4547.
- (51) Moysan, E.; Bastiat, G.; Benoit, J. P. *Gemcitabine versus Modified Gemcitabine: A Review of Several Promising Chemical Modifications* **Mol. Pharm.** (2013), 10, 430.
- (52) Konerding, D.; James, T. L.; Trump, E.; Soto, A. M.; Marky, L. A.; Gmeiner, W. H. *NMR structure of a gemcitabine-substituted model Okazaki fragment* **Biochemistry** (2002), 41, 839.
- (53) Peacock, H.; Kannan, A.; Beal, P. A.; Burrows, C. J. *Chemical Modification of siRNA Bases To Probe and Enhance RNA Interference* **J. Org. Chem.** (2011), 76, 7295.
- (54) Egli, M. *Conformational preorganization, hydration, and nucleic acid duplex stability* **Antisense Nucleic Acid Drug Dev.** (1998), 8, 123.
- (55) Kool, E. T. *Preorganization of DNA: Design principles for improving nucleic acid recognition by synthetic oligonucleotides* **Chem. Rev.** (1997), 97, 1473.

Capítulo 4

Determinación estructural de ácidos nucleicos por RMN

4.1. La RMN: herramienta para la determinación estructural de ácidos nucleicos

Aunque la Resonancia Magnética Nuclear (RMN) es una técnica tradicionalmente utilizada para el análisis de compuestos orgánicos, desde mediados de la década de los 80 ha destacado como técnica para la determinación estructural de macromoléculas biológicas. Richard Ernest fue el primero en proponer y desarrollar las bases teóricas de la RMN actual y los experimentos bidimensionales, ganando por ello el premio Nobel de Química en 1991. A partir de entonces, la expansión de esta técnica ha sido fundamental en el desarrollo tecnológico de multitud de ámbitos. Entre las aplicaciones biofísicas de la técnica hay que destacar la determinación estructural de proteínas y ácidos nucleicos, cuya base metodológica fue establecida por Kurt Wütrich (premio Nobel de Química en 2002).

La importancia de la RMN deriva de la información estructural y dinámica que se puede extraer de las biomoléculas en disolución. Esta técnica es capaz de resolver, con resolución atómica, estructuras macromoleculares en condiciones similares a las observadas en la célula, lo que permiten extraer información dinámica de la estructura. Es en este último aspecto, donde la RMN destaca sobre el resto de técnicas biofísicas empleadas actualmente. Esta técnica permite determinar una gran variedad de fenómenos en muy diversas escalas de tiempo, que van desde los picosegundos a los segundos. Esta propiedad permite que se puedan medir cambios conformacionales y químicos de las moléculas de estudio.

La información dinámica es especialmente importante en ácidos nucleicos, dado que suelen presentar regiones de alta flexibilidad conformacional. Estas regiones, a su vez, suelen estar involucradas en los procesos de reconocimiento molecular del DNA y RNA, en los cuales la RMN es una herramienta perfecta para caracterizarlos.

El desarrollo de distintos avances técnicos hace posible la determinación estructural de fragmentos de DNA y RNA cada vez de mayor tamaño y complejidad¹. En este sentido hay que destacar la utilización de imanes

superconductores de alto campo (frecuencia de resonancia de 800-1000 MHz para el ^1H), la introducción de la espectroscopia bi- y multidimensional y los avances en el marcaje isotópico de moléculas de DNA y RNA^{2,3}.

4.1.1 Fundamentos físicos de la RMN

La técnica de RMN se utiliza con núcleos que poseen espín, también denominados núcleos magnéticamente activos, que se comportan como pequeños imanes. En ausencia de campo magnético, los espines se orientan al azar. Sin embargo, cuando una muestra se coloca en un campo magnético estático, B_0 , los espines tienden a orientarse y aparece una magnetización macroscópica neta en la dirección de B_0 (magnetización longitudinal). Los núcleos con espín $\frac{1}{2}$ pueden orientarse en la misma dirección o en dirección opuesta al campo magnético. Las dos direcciones se corresponden con estados de diferente energía, denominados α y β . En el equilibrio existen más núcleos en el estado α , de menor energía, que en el estado β . La particularidad de la espectroscopía de Resonancia Magnética Nuclear con respecto a otras técnicas, se basa principalmente en que la diferencia de energía entre los dos estados es muy pequeña. Este efecto tiene dos consecuencias principales: (I) que son necesarias frecuencias de resonancia muy bajas, en el rango de las radiofrecuencias, y (II) que la diferencia de poblaciones en los dos estados no puede ser muy grande. Esta pequeña variación de poblaciones en el estado α y β tiene como consecuencia que la técnica sea muy poco sensible en comparación con otras espectroscopias. En la práctica, este efecto provoca que sea necesario trabajar con muestras muy concentradas, típicamente en el rango milimolar.

La diferencia de energía entre los dos estados de espín, α y β , depende de la intensidad del campo magnético aplicado. Cuanto mayor sea el campo magnético, mayor diferencia energética habrá entre los dos estados de espín, y por tanto la diferencia de poblaciones entre ambos será mayor. Este efecto tiene como consecuencia que la sensibilidad de la técnica aumente a campos magnéticos más intensos. A lo largo de esta tesis hemos utilizado espectrómetros de 600, 700 y 800 MHz. A continuación se describen brevemente los experimentos de RMN que se han utilizado y la información estructural que se extrae de ellos.

4.1.2. RMN bidimensional

En los espectros monodimensionales (1D) de RMN de macromoléculas existe generalmente un alto grado de solapamiento entre las distintas señales debido al gran número de protones presentes en la muestra. El análisis de los espectros resulta complicado y, para poder asignar por completo todas las resonancias de la molécula, se hace necesario recurrir a la RMN bidimensional (2D).

La secuencia general de pulsos para un espectro bidimensional es la que se muestra de forma esquemática en la Figura 1. La variación de la señal durante el tiempo de adquisición (t_2) es lo que se conoce como FID, y su transformada de Fourier da lugar a una señal de frecuencias característica. Si se realizan numerosos experimentos del mismo tiempo, variando para cada uno de ellos el tiempo de evolución (t_1), la variación de la magnetización con el tiempo t_1 , para cada valor del tiempo t_2 , origina una curva similar a la FID sobre la que también se puede realizar una transformada de Fourier. El resultado de un experimento de estas características consiste en una matriz de datos, donde las filas contienen información variable con t_2 a un valor fijo de t_1 , y las columnas contienen información variable de t_1 a un valor fijo de t_2 . La transformada de Fourier en ambos dominios de tiempo, t_1 y t_2 , origina el espectro bidimensional en el que cada señal es función de dos frecuencias, ω_1 en la dimensión vertical y ω_2 en la horizontal. Las señales pertenecientes a la diagonal del espectro bidimensional o 2D homonuclear corresponden al espectro 1D, mientras que las señales fuera de ella correlacionan los núcleos entre los que se ha producido transferencia de magnetización durante el tiempo de mezcla τ_m .

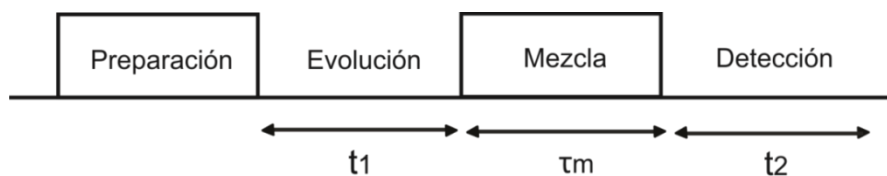


Figura 1: Secuencia de pulsos general para la obtención de un espectro bidimensional.

Dependiendo del tipo concreto de experimento bidimensional que se quiere realizar, puede eliminarse parcialmente el solapamiento y, además, se puede obtener información estructural que dependerá del fenómeno que relacione los valores de frecuencias en ambas dimensiones.

La secuencia general de pulsos consta, en primer lugar, de un periodo de preparación, en el que el sistema sale del equilibrio y, típicamente, se convierte la magnetización longitudinal en transversal. A continuación, hay un periodo de evolución en el que se registra la información sobre una frecuencia que evoluciona durante t_1 . Durante el tiempo de mezcla, el sistema parte de un estado de no equilibrio y se produce una transferencia de magnetización del núcleo cuya frecuencia evolucionó con t_1 a un núcleo con el que esté conectado, escalar o dipolarmente, y cuya frecuencia evolucionará durante el tiempo de mezcla de t_2 . En función de cómo sea este periodo de mezcla se obtendrán los diferentes experimentos bidimensionales entre dos núcleos de distinta frecuencia. El resultado será un pico de correlación en el espectro bidimensional entre dos protones de distinta frecuencia. Si en la Figura 1 se introducen un mayor número de periodos de evolución y de mezcla, se obtienen los espectros tri- (3D) y tetradimensionales (4D), que pueden ser necesarios para la determinación estructural de biomoléculas con elevado peso molecular ($\sim 25\text{kD}$).

La transferencia de magnetización entre los núcleos en el tiempo de mezcla se produce entre átomos con spin $\frac{1}{2}$. El átomo más abundante en la naturaleza que posee esta característica es el protón (^1H). Por tanto, los experimentos bidimensionales más utilizados en la determinación estructural de ácidos nucleicos se denominan homonucleares y correlacionan protones entre sí. Entre los más destacados encontramos los experimentos COSY, TOCSY y NOESY, que explicaremos a continuación y que se han utilizado en esta tesis. Por otro lado, los experimentos denominados heteronucleares correlacionan diferentes tipos de átomos con spin $\frac{1}{2}$. Por ejemplo, ^{13}C , ^{15}N , ^{19}F o ^{31}P . Dado que este trabajo se basa en la determinación estructural de ácidos nucleicos modificados, principalmente con átomos de flúor, también detallaremos los experimentos

bidimensionales HETCOR y HOESY, para correlacionar protones (^1H) con átomos de flúor (^{19}F).

4.1.2.1. Experimentos homonucleares ^1H - ^1H

Experimento COSY (CORelated Spectroscopy)

Las señales de correlación observadas en el espectro COSY⁴ conectan protones que se encuentran acoplados escalarmente, ya que la transferencia de magnetización se produce a través de los electrones de los enlaces. Para que la correlación sea efectiva, ha de tratarse de constantes de acoplamiento a dos y tres enlaces. La secuencia de pulsos del experimento COSY (Figura 2) es muy sencilla y consiste en aplicar dos pulsos no selectivos de radiofrecuencias separados por un tiempo variable denominado t_1 . La adquisición se produce después del segundo pulso durante el periodo t_2 .

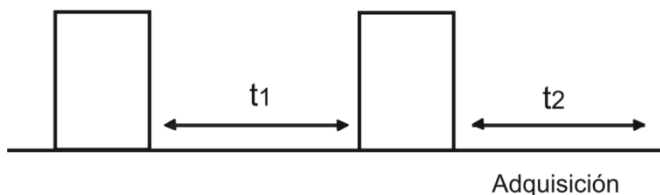


Figura 2: Secuencia de pulsos del experimento COSY. Los rectángulos representan pulsos de 90° .

En el espectro obtenido, las señales de correlación están en modo de absorción y las de la diagonal de dispersión, pero todas ellas son señales en antifase. Esta forma antifase de las señales provoca una cancelación parcial del centro del multiplete, de manera que se incrementa la separación de los componentes dando lugar a una constante de acoplamiento mayor de lo que en realidad es. Para solucionar este inconveniente existen varios experimentos tipo COSY diseñados para determinar cuantitativamente valores de la constante de acoplamiento entre los protones. Entre ellos encontramos el DQF-COSY⁵ o el E-COSY⁶.

A lo largo de esta tesis se ha utilizado el experimento DQF-COSY (*Double Quantum filtered COSY*) para determinar constantes de acoplamiento. En este tipo de experimentos se solucionan los problemas de fase del COSY. Los picos de la

diagonal y los de correlación se encuentran en antifase, lo que provoca que la cancelación afecte de la misma forma a ambos y disminuya la diferencia de intensidad entre ellos.

Experimento TOCSY (Total Correlation Spectroscopy)

El experimento TOCSY⁷ está basado en la condición de Hartmann-Hahn, que establece que la transferencia neta de magnetización entre dos espines acoplados escalarmente es posible si sus frecuencias de precesión efectivas son iguales. Para que se produzca la transferencia de Hartmann-Hahn debe cumplirse que la diferencia entre los campos efectivos experimentados por los dos espines acoplados escalarmente sea menor que la constante de acoplamiento entre ellos. Conseguir con exactitud esta condición es muy difícil ya que, para que esto ocurra, los espines acoplados escalarmente deben situarse equidistantes al transmisor. Sin embargo, normalmente los espines acoplados tienen diferente desplazamiento químico al sufrir un campo magnético efectivo distinto⁸.

Para solucionar este efecto, es posible transferir magnetización entre todos los espines pertenecientes a una misma red de acoplamientos, bloqueando los espines con un campo intenso de radiofrecuencias. Este tipo de secuencia (Figura 3) es capaz de eliminar los términos de desplazamiento químico en anchuras de varios KHz.

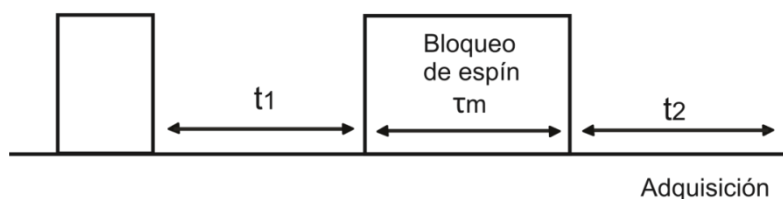


Figura 3: Secuencia de pulsos de un experimento TOCSY. El rectángulo pequeño simboliza el pulso de 90° y el grande un tren de pulsos para bloquear la magnetización.

La secuencia de pulsos de un experimento TOCSY consta en primer lugar de un pulso de 90° en el que la magnetización longitudinal se transforma en transversal. Esta magnetización evoluciona en un tiempo t_1 y, mediante la aplicación de

campos de radiofrecuencias de elevada potencia, se consigue el bloqueo de la magnetización transversal. Este bloqueo de espines permite que se cumpla la condición de Hartmann-Hahn, ya que no hay evolución con los desplazamientos químicos, que permite alcanzar las condiciones de mezcla isotrópica. Las secuencias de pulsos más utilizadas son MLEV16⁹ y MLEV17¹⁰.

En un espectro TOCSY tanto las señales de la diagonal, como los picos de correlación, son señales de absorción y se encuentran en fase. Estos espectros correlacionan un determinado protón con todos los pertenecientes a su sistema de espín. Sin embargo, no permiten distinguir entre las correlaciones directas e indirectas. Esto provoca que su análisis deba hacerse conjuntamente con el del espectro COSY.

Experimento NOESY (Nuclear Overhauser Effect Spectroscopy)

El efecto Overhauser (NOE) se basa en mecanismos de relajación nuclear cruzada. Se manifiesta en un cambio de intensidad de una señal de resonancia cuando las transiciones de otra han sido perturbadas de alguna manera. Depende de la interacción entre dipolos magnéticos y es, tanto más efectiva, cuanto menor es la distancia entre ellos. La distancia entre espines es un factor importante del que depende la velocidad de crecimiento del NOE¹¹. Si se utilizan tiempos de mezcla cortos, es posible aplicar una aproximación en la que la intensidad del NOE se relaciona con la inversa de la sexta potencia de la distancia. Esta fuerte dependencia con la distancia provoca que solo se observen NOEs entre protones cercanos en el espacio ($\sim 6\text{\AA}$).

La secuencia de pulsos del experimento NOESY¹²⁻¹⁴ es la que se muestra en la Figura 4. El primer pulso aplicado genera magnetización transversal que evoluciona libremente durante el periodo de t_1 . A continuación, el segundo pulso convierte una de las componentes de esta magnetización en magnetización longitudinal de no equilibrio. Durante el tiempo de mezcla, el sistema vuelve progresivamente al equilibrio mediante procesos de relajación dipolar, dándose transferencia de magnetización entre los núcleos próximos en el espacio. Finalmente, el último pulso transforma la magnetización transferida al plano xy que se detecta durante

t_2 . La doble transformada de Fourier, en ambos dominios de tiempo, da lugar a un espectro bidimensional donde aparecen los picos de correlación entre núcleos.

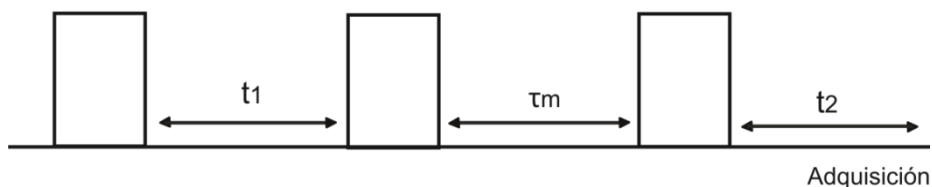


Figura 4: Secuencia de pulsos del experimento NOESY. Los rectángulos simbolizan pulsos de 90° .

Las señales de un espectro NOESY se encuentran en modo absorción pura y en fase. Su intensidad, por tanto, está relacionada con: (I) la velocidad del proceso de relajación de los núcleos, (II) la duración del tiempo de mezcla, y (III) el tiempo de correlación molecular (τ_c). El espectro NOESY, por tanto, está basado en NOEs transitorios.

El experimento NOESY, en algunas ocasiones, representa un problema a la hora de analizar los espectros. Por un lado, la transferencia de magnetización se puede dar a través de dos mecanismos: relajación cruzada o intercambio químico. Cuando ambos procesos son posibles, el experimento NOESY no permite distinguir el origen de los picos de correlación. Por otro lado, en los espectros NOESY pueden aparecer picos de cruce por difusión de espín. La difusión de espín hace referencia a correlaciones entre dos núcleos que pueden no encontrarse próximos en el espacio, pero sí cerca de un tercero. Estos problemas se solucionan con el experimento ROESY¹⁵ (*Rotating Frame nuclear Overhauser Effect Spectroscopy*).

4.1.2.2. Experimentos heteronucleares ^1H - ^{19}F

Experimento HETCOR (HETeronuclear CORrelation)

El experimento HETCOR es muy similar al COSY, lo que permite conectar átomos que se encuentran acoplados escalarmente (Figura 5). La única diferencia es que en este tipo de experimento se correlacionan dos núcleos diferentes. Se aplican

dos pulsos de 90° , el primero, sobre el protón y lo que hace evolucionar la señal con el desplazamiento químico en el tiempo t_1 . El segundo pulso se aplica simultáneamente a los dos tipos de núcleos y permite la transferencia de magnetización del protón al heteronúcleo a través del acoplamiento escalar entre ambos. La intensidad de la señal se modula a través del desplazamiento químico de heteronúcleo, y se adquiere en el tiempo t_2 .

Las propiedades magnéticas del núcleo de ^{19}F hace que este tipo de experimentos puedan realizarse tanto con detección en ^1H como en ^{19}F , al contrario de lo que ocurre con otros heteronúcleos donde la detección en ^1H es siempre más conveniente.

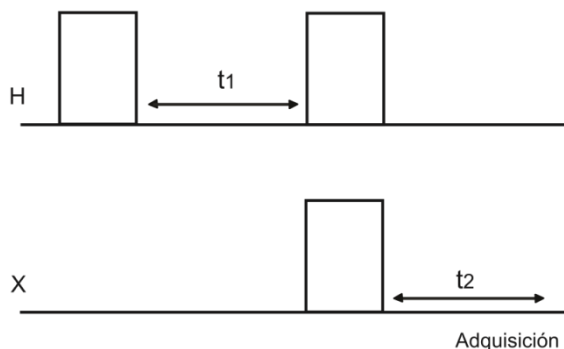


Figura 5: Secuencia de pulsos del experimento HETCOR. Los rectángulos simbolizan pulsos de 90° y la X el heteroátomo, en nuestro caso ^{19}F .

Experimento HOESY (Heteronuclear Overhauser Enhancement Spectroscopy)

El experimento HOESY¹⁶ permite detectar conectividades tipo NOE entre heteroátomos. Las bases de la secuencia de pulsos de este experimento son muy similares a las del experimento NOESY. El primer pulso aplicado genera magnetización transversal, que evoluciona de acuerdo con la constante de acoplamiento con el heteroátomo durante el periodo de t_1 . A continuación, el segundo pulso convierte una de las componentes de esta magnetización en magnetización longitudinal de no equilibrio. Durante el tiempo de mezcla, el sistema vuelve al equilibrio mediante procesos de relajación dipolar, dándose transferencia de magnetización entre los núcleos próximos en el espacio.

Finalmente, el último pulso se aplica sobre el heteroátomo, lo que crea una magnetización transversal que se detecta durante t_2 . La señal detectada se desacopla a protón y, mediante la doble transformada de Fourier, se obtiene un espectro bidimensional, donde aparecen los picos de correlación entre heteronúcleos cercanos en el espacio.

4.1.3. Información estructural extraída de los experimentos de RMN

De los experimentos mono- y bidimensionales de RMN se puede extraer la información necesaria para determinar la estructura tridimensional de la molécula en disolución. Dependiendo del espectro se podrá obtener diferentes tipos de información. De todos ellos, la información más útil en el cálculo de estructuras proviene de las distancias entre átomos obtenidas de los picos de correlación dipolar (NOE). Esta información se complementa, además, con restricciones angulares derivadas del análisis de las constantes de acoplamiento escalar a tres enlaces obtenidas del experimento DQF-COSY⁵.

4.1.3.1. Desplazamiento químico

El campo magnético neto experimentado por el núcleo no es exactamente el campo externo aplicado B_0 , ya que éste se va modificando por el efecto de la nube electrónica que rodea al núcleo. Los electrones que se encuentran alrededor del núcleo inducen campos magnéticos que contrarrestan este campo externo B_0 . Este fenómeno es conocido como apantallamiento. Por tanto, el campo efectivo experimentado de un núcleo determinado viene dado por la ecuación:

$$B=B_0 (1-\sigma)$$

siendo B el campo efectivo experimentado y σ la constante de apantallamiento. Puesto que los diferentes núcleos de una misma especie se encuentran en diferentes entornos en función de la posición que ocupen en la estructura molecular que los contiene, se deberían encontrar tantas frecuencias de resonancia como entornos químicos diferentes haya en esos núcleos de la molécula de estudio.

La aplicabilidad de la RMN a la elucidación estructural y a la identificación de compuestos, surge de estas correlaciones entre las frecuencias de precesión de los núcleos magnéticos y el entorno químico de los mismos. Las frecuencias de precesión varían con la intensidad del campo magnético bajo el que se lleve a cabo el experimento de RMN, sin embargo, el porcentaje de desplazamiento se mantiene cualquiera que sea el valor de B_0 .

Así, se conoce con el nombre de desplazamiento químico δ , al cambio o desplazamiento de la frecuencia de resonancia de un núcleo como consecuencia de la estructura química en la que se encuentre formando parte. Se expresa según la ecuación:

$$\delta = \frac{(\nu_N - \nu_{ref})10^6}{\nu_{ref}}$$

donde ν_N es la frecuencia del núcleo N sujeto a estudio y ν_{ref} es la frecuencia de resonancia de un núcleo elegido como referencia. Las unidades de desplazamiento químico, δ , son partes por millón (ppm).

Al ser el valor de δ extremadamente sensible al entorno químico del núcleo, éste también variará con la conformación adoptada por la molécula estudiada. Sin embargo, como consecuencia de esta gran sensibilidad, la relación entre desplazamientos químicos y los parámetros geométricos y estructurales de la molécula no se pueden formular con la suficiente precisión. Se han hecho algunos intentos al respecto a partir de estudios estadísticos en biomoléculas conocidas. En el caso de ácidos nucleicos se han obtenido resultados prometedores con estructuras canónicas de DNA de doble hélice¹⁷.

Las resonancias correspondientes a los protones imino de las bases nitrogenadas se desplazan hacia campos más bajos cuando se forma el enlace de hidrógeno. Para cada tipo de apareamiento los desplazamientos químicos varían, como ocurre para los pares de bases tipo Hoogsteen, donde las señales correspondientes a los iminos aparecen a campos mayores (9.5-10 ppm) que en el caso de pares de bases tipo Watson y Crick (12-15 ppm). De este modo, los

desplazamientos químicos son un claro indicador del tipo de estructura secundaria del oligonucleótido de estudio. El desplazamiento químico es el parámetro más sensible de los obtenidos por RMN, por esta razón, pequeñas variaciones en éste permiten determinar alteraciones en la estructura de la biomolécula a estudio.

4.1.3.2. NOE

Como ya se ha comentado anteriormente, las restricciones de distancia se obtienen a partir de la intensidad de los picos NOE. Ésta depende de la distancia entre protones y del grado de movilidad de la molécula. Una primera aproximación para calcular las distancias interprotónicas es la llamada aproximación de dos espines, dónde la intensidad del NOE se considera proporcional al tiempo de mezcla del experimento NOESY y decrece con la sexta potencia de la distancia interprotónica, tal y cómo se indica en la expresión:

$$NOE \propto \frac{\tau_m}{r^6}$$

Esta aproximación supone que la molécula se mueve como un sólido rígido y que cada par de protones se encuentra aislado del resto. Para calcular la constante de proporcionalidad se utiliza la intensidad del NOE entre protones, cuya distancia es independiente de la conformación, como ocurre con los protones geminales y aromáticos. Esta expresión presenta varias limitaciones. En primer lugar, sólo es válida para tiempos de mezcla cortos, en los que el efecto de difusión de espín es despreciable. En segundo lugar, la suposición de movimiento isotrópico en el que la molécula se mueve cómo un sólido rígido con un único tiempo de correlación global, no es del todo exacta en el caso de macromoléculas, ya que generalmente éstas presentan un comportamiento dinámico muy complejo, donde tienen lugar movimientos internos en varias escalas de tiempo. En el caso de que estos movimientos internos sean rápidos en comparación con el movimiento global de la molécula, se puede utilizar el formalismo de Lipari y Szabo para correlacionar el efecto NOE con la geometría molecular y un parámetro denominado de orden, que contiene toda la información dinámica local¹⁸. Un caso especial de movimiento interno rápido es la rotación de los tres protones de los grupos metilo. En este caso en particular, se llega a una expresión parecida a la anterior, donde la

intensidad del NOE es proporcional a la inversa de la tercera potencia de la distancia interprotónica¹¹:

$$NOE \propto \frac{\tau_m}{r^3}$$

En la práctica, la validez de estas expresiones para un caso particular es muy difícil de asegurar. En muchos casos, se clasifica cualitativamente la intensidad del NOE en fuerte medio y débil, y se hace corresponder una restricción de distancia, que usualmente se consideran entre 1.8-3.0, 1.8-4.0 y 1.8-5.0 Å, respectivamente.

En el caso de ácidos nucleicos es necesario obtener restricciones de distancias muy precisas. La obtención de restricciones de distancias precisas se consigue aplicando la teoría de la matriz completa de relajación¹⁹. Las intensidades de los picos de cruce (NOEs) para un tiempo de mezcla $V(\tau_m)$, en un espectro NOESY bidimensional, pueden ser recalculadas a partir de las intensidades de los picos de la diagonal, $V(0)$, a un tiempo de mezcla $\tau_m=0$. Estos parámetros se relacionan con la geometría molecular mediante las expresiones:

$$V(\tau_m) = \exp(-R\tau_m) V(0)$$

donde τ_m es el tiempo de mezcla del experimento, V es la matriz de intensidades NOE y R representa la matriz relajación.

Aunque utilizando esta aproximación se mejora considerablemente el cálculo de restricciones experimentales, los valores finales no están exentos de errores. En la práctica, para minimizar este problema, en el cálculo inicial se utilizan distintos modelos de partida como son los DNA canónicos tipo A y B con diferentes tiempos de correlación y de mezcla.

La restricción final que se obtiene no es una distancia determinada entre dos protones, sino un rango de distancias con una cota inferior y otra superior correspondientes a los errores en la determinación de dicha distancia. En el caso de utilizar métodos de matriz relajación, este rango resulta más estrecho y las estructuras estarán mejor definidas.

A la hora de determinar la estructura tridimensional de los ácidos nucleicos, no todas las distancias tienen la misma relevancia. Existen muchas distancias que no dependen de la conformación, y que por lo tanto, no son útiles. Sin embargo, otras son de gran importancia, como ocurre con las distancias interprotónicas entre los átomos $d_i(2'',4')$ y $d_i(1',4')$, que determinan la conformación del azúcar (Figura 6)²⁰. El ángulo glicosídico también viene dado por la distancia $d_i(1',6/8)$, que establece si la base se encuentra en posición *syn* o *anti* con respecto al azúcar (capítulo 2). Las distancias secuenciales o intercadena son todas importantes ya que dan información del grado de estructuración de la molécula y del tipo de estructura que forman (tipo A o B).

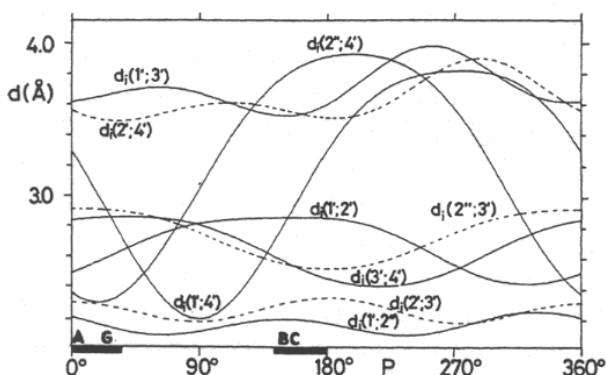


Figura 6: Variación de las distancias entre los diferentes protones del anillo de desoxiribosa en función del ángulo de pseudorotación P . Sobre el eje se ha indicado la fase del ángulo de pseudorotación preferido para las formas A- y B-ADN, así como para los nucleótidos C y G en las formas Z-ADN en el oligonucleótido $d(CGCGCG)_2$.

4.1.3.3 Constantes de acoplamiento (J)

Otra de las magnitudes observables mediante RMN es el llamado acoplamiento escalar. Es una interacción magnética que se transmite por los electrones de enlace. Este fenómeno provoca que la frecuencia de resonancia de un núcleo se vea alterada por la presencia de otro, u otros núcleos magnéticos, que se encuentran directamente enlazados al primero, o distantes por un número limitado de enlaces. Esta interacción da como resultado un desdoblamiento de las señales en el espectro.

El acoplamiento de espín se relaciona con un parámetro espectral denominado acoplamiento escalar, J que normalmente se extrae de los experimentos tipo COSY. Desde el punto de vista estructural, la constante de acoplamiento más importante es la constante a tres enlaces. Esta constante depende del ángulo diedro definido por los tres enlaces, a través de la denominada ecuación de Karplus:

$$J_{HH} = P_1 \cos^2 \varphi + P_2 \cos \varphi + P_3 + \sum_1 \Delta\chi_i [P_4 + P_5 \cos^2 (\xi_i \varphi + P_6 |\Delta\chi_i|)] - \Delta J$$

donde φ es el ángulo de torsión del frgménto H-C-C-H; el sumatorio se hace sobre todos los α -sustituyentes de los carbonos; ξ toma el valor de +1 o -1 dependiendo de la orientación positiva o negativa del sustituyente y con respecto al protón geminal²¹; $\Delta\chi_i$ representa la diferencia de electronegatividad²² entre los α -sustituyentes y el hidrógeno; ΔJ tiene en cuenta la interacción espín-espín a través del espacio²³. Los parámetros empíricos P_1 , P_2 , P_3 , P_4 , P_5 , P_6 y P_7 dependen del número de sustituyentes en α . Habitualmente se usa la ecuación de Karplus de forma simplificada:

$$^3J(\theta) = A \cos^2 \theta + B \cos \theta + C$$

donde θ es el ángulo de torsión.

Una de las aplicaciones más habituales de las constantes de acoplamiento vecinales protón-protón en el campo de los ácidos nucleicos es la determinación de la conformación del azúcar. Como ya se ha comentado anteriormente, algunas correlaciones NOE entre protones del azúcar permiten extraer información de la conformación de éste. Sin embargo, estas medidas muchas veces no son suficientes para dar una descripción precisa de la conformación del azúcar. Afortunadamente, se puede obtener información mucho más precisa a través de las constantes de acoplamiento.

Dado que los ángulos de torsión protón-protón están relacionados con los ángulos de torsión endocíclicos, se puede establecer una relación entre las constantes de

acoplamiento individuales y los parámetros de pseudorrotación, el ángulo de fase (P) y la amplitud de ángulo de pseudorrotación (v_m) (ver capítulo 2).

De este modo se obtiene un conjunto de constantes de acoplamiento para diferentes valores de P y v_m . Puesto que podemos medir cinco constantes entre distintos tipos de protones en el azúcar, y la conformación del azúcar se describe con únicamente dos parámetros, el problema conformacional está sobredeterminado. No obstante, es muy habitual que no todas las constantes de acoplamiento se puedan ajustar a un valor de P y v_m . La razón de ello es la existencia de un equilibrio dinámico entre varias conformaciones del anillo. En esta situación el valor de la constante de acoplamiento será la suma de las contribuciones de los diversos conformeros²⁴.

4.1.3.4. Cálculo de las constantes de acoplamiento de nucleótidos modificados en C2'

La constante de acoplamiento entre protones (J_{HH}), para cada conformación del azúcar, depende de la diferencia de electronegatividad entre el hidrógeno y los α -sustituyentes ($\Delta\chi_i$). Dado que la mayoría de los nucleótidos utilizados en esta tesis contienen átomos de flúor en el carbono C2', las constantes de acoplamiento (J_{HH}) de éstos serán distintas a las del DNA y RNA. Por ello, se han recalculado los valores de J_{HH} de 2'F-ANA, 2'F-RNA y ANA para cada P y $v_m=35^\circ$, de acuerdo con la ecuación de Karplus. En las figuras 7 y 8 se presentan las constantes de acoplamiento más relevantes para la determinación de la estructura del azúcar.

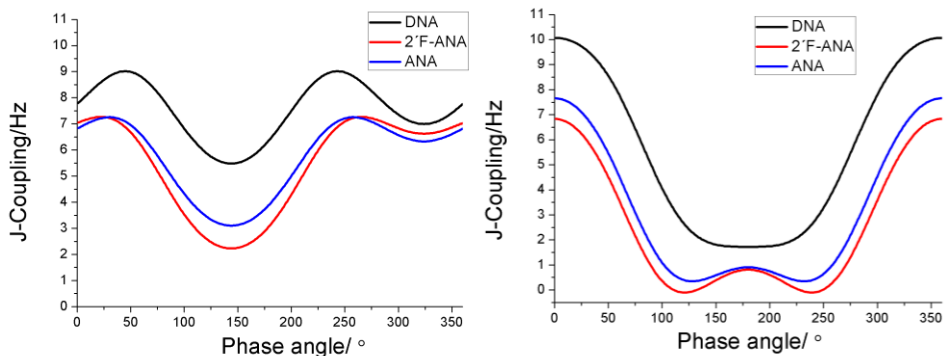


Figura 7: DNA en negro, 2'-F-ANA en rojo y ANA en azul: $J_{1'2''}$ (izquierda) y $J_{2'3'}$ (derecha).

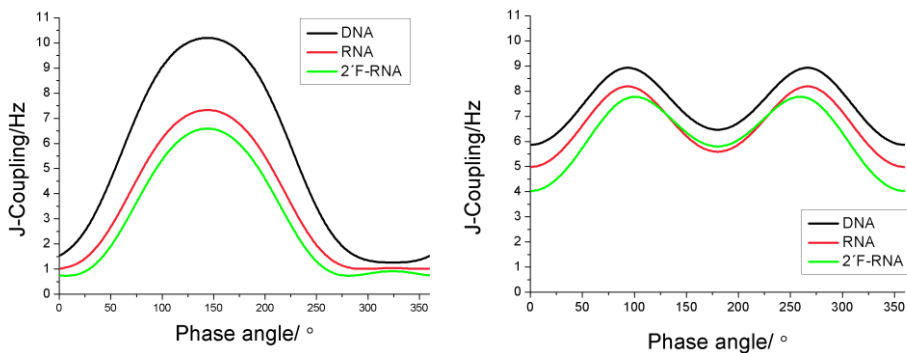


Figura 8: DNA en negro, RNA en rojo y 2'-F-RNA en verde: $J_{1'2'}$ (izquierda) y $J_{2'3'}$ (derecha).

4.2. Cálculo de estructuras en disolución

La determinación estructural de una biomolécula en disolución consiste en obtener las coordenadas atómicas compatibles con las restricciones experimentales de distancias interprotónicas y ángulos diedros derivados de la RMN. No se trata de encontrar la estructura que mejor cumpla estas restricciones experimentales, sino representar todo el conjunto de estructuras compatibles con la información

experimental. Por este motivo, las estructuras determinadas por RMN se muestran como una superposición de varias conformaciones.

Para conseguir este fin, se han desarrollado diversos algoritmos de cálculo para ajustar toda esta información experimental y a la vez explorar adecuadamente el espacio conformacional. Las diferentes metodologías de cálculo se clasifican en función de la aproximación matemática que utilizan. En esta tesis, se han utilizado fundamentalmente técnicas basadas en la Dinámica Molecular.

4.2.1 Métodos de simulación

Existen diferentes tipos de técnicas teóricas aplicables basadas en la mecánica clásica. Las más importantes son: la mecánica molecular, la dinámica molecular, los métodos de Monte Carlo y la mecánica estadística. En esta sección tan solo se explicarán las dos primeras técnicas que son las que emplearemos en el refinamiento de estructuras de ácidos nucleicos.

4.2.1.1. Minimización de energía

La técnica clásica más sencilla es la minimización de energía²⁵. Dado un conjunto de variables independientes, como son las coordenadas atómicas, y una función de energía potencial, el algoritmo de minimización de la energía consiste en encontrar un conjunto de valores para estas coordenadas que hagan mínimo el valor de la función potencial.

Desde el punto de vista computacional, el problema de la minimización de energía de un sistema modelo macromolecular es encontrar el mínimo de una función general no lineal. Los dos algoritmos más comúnmente utilizados son el “*steepest descent*”, que consiste en buscar en cada iteración el mínimo de la función en la dirección indicada por el gradiente (dirección en la que el descenso es más pronunciado); y el método de gradiente conjugado, en el que la dirección de búsqueda se basa en encontrar una dirección resultante de combinar la dirección del gradiente durante las últimas iteraciones.

Estos métodos se utilizan con un algoritmo iterativo que realiza los siguientes pasos: (I) evalúa la energía para una configuración dada; (II) determina la primera

derivada de la energía con respecto a las coordenadas; (III) estima la dirección de búsqueda del mínimo y se obtienen unas nuevas coordenadas; (IV) evalúa la energía de la nueva configuración, y si ésta converge, el cálculo se da por finalizado. Si no se produce la convergencia el proceso se repite hasta que esto ocurra.

En el caso de macromoléculas de DNA o RNA, al tener un número muy amplio de grados de libertad, tan solo se puede explorar una fracción muy pequeña de todo el espacio conformacional. Por tanto, los métodos de minimización de energías se usan para refinar estructuras moleculares, para eliminar los contactos estéricos desfavorables, y ajustar las longitudes y ángulos de enlace a valores óptimos. También pueden dar información sobre las energías relativas de diferentes conformaciones, aunque generalmente no van a generar estructuras muy distintas a la inicial.

4.2.1.2 Dinámica Molecular

La dinámica molecular²⁶⁻²⁹ es una herramienta muy poderosa y utilizada en una gran variedad de áreas de la ciencia. El objetivo de las simulaciones es estudiar las propiedades estructurales, termodinámicas y dinámicas de un sistema molecular. Es por ello que esta técnica aporta información sobre la dependencia del tiempo y la magnitud de las fluctuaciones tanto en posiciones como en velocidades de las partículas del sistema. La dinámica molecular considera a cada átomo como una masa puntual cuyo movimiento está determinado por las fuerzas que ejercen sobre él los demás átomos. Esta aproximación es válida si se trabaja a temperatura ambiente y si el proceso estudiado no implica la formación ni la ruptura de enlaces covalentes.

El método de dinámica molecular consiste en integrar las ecuaciones de movimiento de Newton para cada átomo del sistema. En un sistema de N átomos, la fuerza ejercida sobre el átomo i en un instante de tiempo t , viene dada por un gradiente negativo de energía potencial, $V(r_1, r_2, \dots, r_N)$:

$$F_i(t) = -\frac{\partial}{\partial r_i} V(r_1, r_2, \dots, r_N)$$

Una vez conocida la fuerza que actúa sobre la partícula i , se puede calcular la aceleración y velocidad que tendrá la partícula al cabo del incremento del tiempo. La posición de la partícula, finalmente se calcula integrando la velocidad con respecto al tiempo. Este proceso es iterativo. Cada nueva posición de la partícula se evalúa a nivel de la energía y las fuerzas que actúan sobre ella en esta nueva posición y se repite el cálculo de manera que obtenemos un conjunto de posiciones y velocidades de cada partícula a lo largo del tiempo que reciben el nombre de trayectoria.

Para comenzar los cálculos de dinámica molecular se requieren dos valores, las coordenadas iniciales, $X(0)$, y un conjunto de valores de velocidades iniciales, $V(0)$. En esta tesis, las coordenadas iniciales se han obtenido de estructuras estándar tipo A y B. Las velocidades iniciales se asignan aleatoriamente para una temperatura T . En la práctica, las velocidades inicialmente se asignan a baja temperatura, y posteriormente se re escalan gradualmente hasta alcanzar la temperatura deseada mediante dinámicas cortas que equilibran el sistema. Este método es muy útil a la hora de refinar estructuras de RMN. A continuación, para asegurar que el sistema es estable y no presenta fluctuaciones, se hace una simulación larga. Este paso puede tomar desde decenas de picosegundos, a cientos de ellos, dependiendo del tamaño del sistema. La mayoría de las simulaciones de biomoléculas se realiza a temperatura y presión constantes. La simulación dinámica es fiable cuando el cálculo alcanza el equilibrio. La trayectoria generada en este caso se guarda y se analiza posteriormente (Figura 9).

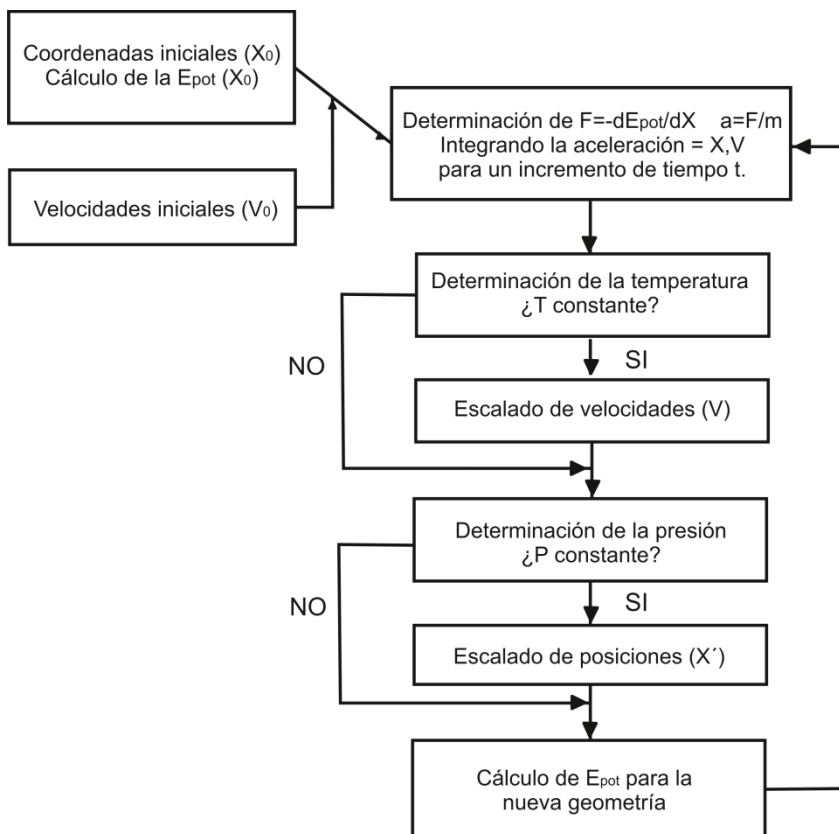


Figura 9: Algoritmo utilizado en los métodos de dinámica molecular.

La principal ventaja de la Dinámica Molecular reside en su capacidad de saltar barreras de potencial y, por tanto, superar el problema de los mínimos locales. Además, si el protocolo de cálculo es adecuado, este método permite explorar eficientemente el espacio conformacional compatible con las restricciones experimentales.

4.2.2. Campo de fuerzas

El campo de fuerzas es un conjunto de expresiones y parámetros que describen las diferentes contribuciones energéticas de las interacciones entre las partículas del sistema. Los campos de fuerzas implementados en los programas AMBER^{30,31}, CHARMM³² y GROMOS^{33,34} están parametrizados para el estudio de macromoléculas biológicas, y disponen de parámetros tanto para proteínas como

para ácidos nucleicos y para complejos proteína-DNA. En esta tesis las simulaciones de dinámica molecular para los diferentes oligonucleótidos estudiados se realizaron fundamentalmente con el programa AMBER 7.0³⁵. Los campos de fuerzas actuales, y utilizados para el cálculo de las estructuras de esta tesis, han sido parametrizados utilizando datos experimentales derivados de diversas técnicas como RMN, IR, difracción de rayos X, etc³⁶. Los diversos campos de fuerzas pueden utilizar diferentes expresiones para evaluar la energía del sistema. En general, viene dado por la contribución de distintos términos evaluados independientemente. Estos se pueden agrupar en dos grandes bloques: los términos enlazantes, correspondientes a átomos enlazados covalentemente, y a los no enlazantes, que hacen referencias a la interacción entre átomos no enlazados covalentemente:

$$E_{total} = E_{enlazantes} + E_{no-enlazantes}$$

La energía de los términos enlazantes engloba un término de tensión de enlace (*stretching*), un término de flexión de ángulo (*bending*), un término de torsión y un término impropio que penaliza la desviación de los átomos fuera de un plano formado por otros tres átomos (Tabla 1). Por otro lado la energía de los términos no enlazantes depende de un término electrostático otro de van der Waals (Tabla 1).

Tabla 1: Expresiones matemáticas típicas, que representan los múltiples términos de energía, tanto enlazantes como no enlazantes.

Términos enlazantes	
Stretching	$E_{str} = K_s(l - l_0)^2$
Bending	$E_{bnd} = K_b(\theta - \theta_0)^2$
Torsión	$E_{tor} = 0.5 \sum_n K_n (1 + \cos(n\psi - B_n))$
Términos no enlazantes	
Electrostático	$E_{ele} = \frac{Q_A Q_B}{\epsilon r}$
Van der Waals	$E_{vW} = m^* [(\frac{r^*}{r})^{12} - 2(\frac{r^*}{r})^6]$

Siendo K_s y K_b las constantes de fuerza correspondientes, l_0 la distancia de equilibrio, l la longitud del enlace químico, θ el ángulo que forman tres átomos y θ_0 el ángulo de equilibrio. El término de torsión se representa como una suma de términos de Fourier de diferente periodicidad, en el que K_n representa la altura de la barrera para cada término, n es la periodicidad y β_n es el ángulo de fase. Para los términos de energía no enlazante, r es la distancia entre dos átomos, r^* la distancia a la cual la energía es mínima y m^* indica la profundidad del pozo atractivo.

4.2.3. Dinámica molecular restringida

Para el cálculo de las estructuras determinadas a lo largo de esta tesis se han utilizado los métodos de dinámica molecular restringida. Para ello se han utilizado datos experimentales de RMN introduciendo un nuevo término de potencial, sin sentido físico, pero que representa dicha información experimental. Este tipo de dinámica molecular recibe el término de dinámica molecular restringida:

$$V = V_{NOE} + V_J + V_\sigma$$

El potencial experimental se puede desglosar en los términos correspondientes a distancias NOE y ángulos de torsión y/o constantes de acoplamiento generalmente de la forma:

$$V_d = \begin{cases} K_{exp}(d - u)^2 & \text{si } d > u \\ 0 & \text{si } l < d < u \\ K_{exp}(d - l)^2 & \text{si } d < l \end{cases}$$

donde u y l representan los límites superior e inferior para dicha magnitud y d representa cualquier magnitud experimental (distancia, ángulo diedro, etc.). El valor de la magnitud es 0 cuando la restricción experimental está comprendida entre u y l , y aumenta progresivamente a medida que nos alejamos de dicho rango.

4.2.4. Parametrización

Los parámetros del campo de fuerzas que aparecen en las expresiones anteriores vienen dados por la geometría molecular, determinada por los términos de distancias, ángulos de enlace en el mínimo de potencial y los ángulos de torsión. Por otro lado, también son necesarios los parámetros determinados a partir del espectro vibracional y las energías conformacionales, obtenidos de las constantes de fuerza y barreras de torsión, respectivamente.

Las distancias, ángulos de enlace y torsiones se pueden obtener de estructuras obtenidas por Rayos X o RMN, sin embargo, los parámetros de estas estructuras generalmente no se corresponden con los de distancia de enlace y ángulos en el mínimo de potencial (valor de equilibrio). Esto es debido a la influencia de otros términos en la función de energía potencial sobre la geometría final. Por tanto, estos parámetros deben calcularse, mediante técnicas cuánticas, directamente para la molécula a estudio. Normalmente, la obtención de los valores de equilibrio se obtiene con relativa fiabilidad mediante cálculos sencillos³⁶. Sin embargo, la obtención de las constantes de fuerza requiere cálculos más sofisticados^{37,38}.

A lo largo de esta tesis ha sido necesario parametrizar los nucleótidos de 2'F-RNA y 2'2'dFdC que contenían grupos funcionales no incluidos en el campo de fuerzas. Los nucleótidos 2'F-ANA y ANA ya se habían parametrizado con anterioridad, lo que nos permitió utilizar los términos correspondientes para el cálculo estructural³⁹. Para obtener las constantes de fuerza, en primer lugar, se varía el término a parametrizar y se calcula a nivel cuántico su energía a cada paso. A continuación, mediante métodos clásicos, se realizan una serie de cálculos similares a los anteriores. Finalmente, la constante de fuerza se varía hasta que el perfil clásico se ajuste lo mejor posible al perfil cuántico.

Un término muy importante es el de las cargas atómicas, que es uno de los pasos más delicados de la parametrización de un campo de fuerzas. Las cargas atómicas dependen en gran medida del entorno molecular y deben ser parametrizadas en cada caso. Aunque existen múltiples estrategias^{37,40} para su obtención, una de las más utilizadas se basa en realizar primero el cálculo del potencial electrostático a nivel mecanocuántico en moléculas modelo sencillas.

Una vez obtenido este potencial, se realiza un ajuste para obtener el conjunto de cargas puntuales que mejor represente el potencial obtenido cuánticamente en la superficie de la molécula⁴¹. Estas cargas, aunque son razonablemente buenas, presentan una serie de inconvenientes. En primer lugar, son muy dependientes de la conformación en las que se han obtenido, lo cual puede inducir errores en la obtención de energías intramoleculares en moléculas flexibles. En segundo lugar, los puntos en los que se calcula el potencial electrostático mecanocuántico se hallan en la superficie de la molécula, y por tanto los átomos que no están expuestos tienen un peso pequeño en el proceso de ajuste. Este efecto ocasiona que los átomos en esta situación tengan cargas artificialmente grandes. Este problema se ha solucionado, en parte, mediante la estrategia RESP⁴² (Restrained ElectroStatic Potential), que consiste en reducir las cargas de estos átomos enterrados mediante una penalización en un proceso de ajuste. RESP, por tanto, da como resultado un conjunto de cargas más seguras y precisas.

BIBLIOGRAFÍA

- (1) Varani, G.; Tinoco, I. *RNA structure and NMR spectroscopy* **Q. Rev. Biophys.** (1991), 24, 479.
- (2) Kim, I.; Lukavsky, P. J.; Puglisi, J. D. *NMR Study of 100 kDa HCV IRES RNA Using Segmental Isotope Labeling* **J. Am. Chem. Soc.** (2002), 124, 9338.
- (3) Werner, M. H.; Gupta, V.; Lambert, L. J.; Nagata, T. *Uniform 13C/15N-labeling of DNA by tandem repeat amplification* **Methods Enzymol.** (2001), 338, 283.
- (4) Aue, W.; Karhan, J.; Ernst, R. *Homonuclear broad band decoupling and two-dimensional J-resolved NMR spectroscopy* **J. Chem. Phys.** (1976), 64, 4226.
- (5) Rance, M.; Sørensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering* **Biochem. Biophys. Res. Commun.** (1983), 117, 479.
- (6) Griesinger, C.; Sørensen, O. W.; Ernst, R. R. *Practical aspects of the E.COSY technique. Measurement of scalar spin-spin coupling constants in peptides* **J. Magn. Reson.** (1987), 75, 474.
- (7) Braunschweiler, L.; Ernst, R. *Coherence transfer by isotropic mixing: application to proton correlation spectroscopy* **J. Magn. Reson.** (1983), 53, 521.
- (8) Hicks, R. P.; Young, J. K.; Moskau, D. *Magnetization transfer via isotropic mixing: an introduction to the HOHAHA experiment* **Concepts Magn. Reson.** (1994), 6, 115.
- (9) Levitt, M. H.; Freeman, R.; Frenkiel, T. *Broadband heteronuclear decoupling* **J. Magn. Reson.** (1982), 47, 328.
- (10) Bax, A.; Davis, D. G. *MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy* **J. Magn. Res.** (1985), 65, 355.
- (11) Neuhaus D., W., M. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, (1989).
- (12) Jeener, J.; Meier, B.; Bachmann, P.; Ernst, R. *Investigation of exchange processes by two-dimensional NMR spectroscopy* **J. Chem. Phys.** (1979), 71, 4546.
- (13) Kumar, A.; Ernst, R.; Wüthrich, K. *A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules* **Biochem. Biophys. Res. Commun.** (1980), 95, 1.

- (14) Macura, S.; Ernst, R. *Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy* **Mol. Phys.** (1980), 41, 95.
- (15) Bax, A.; Davis, D. G. *Practical aspects of two-dimensional transverse NOE spectroscopy* **J. Magn. Res.** (1985), 63, 207.
- (16) Yu, C.; Levy, G. C. *Two-dimensional heteronuclear NOE (HOESY) experiments: investigation of dipolar interactions between heteronuclei and nearby protons* **J. Am. Chem. Soc.** (1984), 106, 6533.
- (17) Wijmenga, S. S.; van Buuren, B. N. *The use of NMR methods for conformational studies of nucleic acids* **Prog. Nucl. Magn. Reson. Spectrosc.** (1998), 32, 287.
- (18) Lipari, G.; Szabo, A. *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity* **J. Am. Chem. Soc.** (1982), 104, 4546.
- (19) Borgias, B. A.; James, T. L. *MARDIGRAS-A procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure* **J. Magn. Reson.** (1990), 87, 475.
- (20) Wüthrich, K. *NMR of Proteins and Nucleic Acids*, (1986).
- (21) Haasnoot, C.; De Leeuw, F.; Altona, C. *The relationship between proton-proton NMR coupling constants and substituent electronegativities—I: an empirical generalization of the Karplus equation* **Tetrahedron** (1980), 36, 2783.
- (22) Huggins, M. L. *Bond Energies and Polarities* **J. Am. Chem. Soc.** (1953), 75, 4123.
- (23) De Leeuw, F. A.; Van Beuzekom, A. A.; Altona, C. *Through-Space effects on vicinal proton spin-spin coupling constants mediated via hetero atoms: Nonequivalence of cis couplings in five-membered rings* **J. Comput. Chem.** (1983), 4, 438.
- (24) Rinkel, L. J.; Altona, C. *Conformational analysis of the deoxyribofuranose ring in DNA by means of sums of proton-proton coupling constants: a graphical method* **J. Biomol. Struct. Dyn.** (1987), 4, 621.
- (25) Van Holde, K. E.; Johnson, W. C.; Ho, P. S. *Principles of physical biochemistry* (2006).
- (26) McCammon, J.; Karplus, M. *Simulation of protein dynamics* **Annu. Rev. Phys. Chem.** (1980), 31, 29.
- (27) McCammon, J. A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press, (1988).
- (28) Cheatham III, T. E.; Kollman, P. A. *Molecular dynamics simulation of nucleic acids* **Annu. Rev. Phys. Chem.** (2000), 51, 435.
- (29) Orozco, M.; Noy, A.; Pérez, A. *Recent advances in the study of nucleic acid flexibility by molecular dynamics* **Curr. Opin. Struct. Biol.** (2008), 18, 185.
- (30) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules* *J. Am. Chem. Soc.* 1995, 117, 5179-5197 **J. Am. Chem. Soc.** (1996), 118, 2309.
- (31) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. *Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers* **Biophys. J.** (2007), 92, 3817.
- (32) MacKerell Jr, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. *An all-atom empirical energy function for the simulation of nucleic acids* **J. Am. Chem. Soc.** (1995), 117, 11946.
- (33) Scott, W. R.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *The GROMOS biomolecular simulation program package* **J. Phys. Chem. A** (1999), 103, 3596.
- (34) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation* **J. Chem. Theory Comput.** (2008), 4, 435.
- (35) David A Case, D. A. P., James W Caldwell, Thomas E Cheatham III, Junmei Wang, Wilson S Ross, CL Simmerling, TA Darden, Kenneth M Merz, Robert V Stanton, AL Cheng, James J Vincent, Mike Crowley, Vickie Tsui, Holger Gohlke, RJ Radmer, Yong Duan, Jed Pitera, Irina Massova, George L Seibel, U Chandra Singh, PK Weiner, Peter A Kollman University of California, San Francisco, 2007.
- (36) Alemán, C.; Bella, J.; Perez, J. J. *Molecular mechanics and molecular dynamics simulations of the hexagonal crystalline form of the helical polyamide poly (α -isobutyl-L-aspartate)* **Polymer** (1994), 35, 2596.
- (37) Bock, C. W.; Panchenko, Y. N.; Pupyshev, V. *Effect of basis set quality and electron correlation on the scale factors of a harmonic force field* **J. Comput. Chem.** (1990), 11, 623.
- (38) Ducéré, J.-M.; Cavallo, L. *Parametrization of an Empirical Correction Term to Density Functional Theory for an Accurate Description of π -Stacking Interactions in Nucleic Acids* **J. Phys. Chem. B** (2007), 111, 13124.
- (39) Noy, A.; Luque, F. J.; Orozco, M. *Theoretical analysis of antisense duplexes: determinants of the RNase H susceptibility* **J. Am. Chem. Soc.** (2008), 130, 3486.
- (40) Fliszar, S. *Charge distributions and chemical effects. XLII. Atomic charges and their role in energy calculations* **Int. J. Quantum Chem** (1986), 29, 305.

- (41) Momany, F. A. *Determination of partial atomic charges from ab initio molecular electrostatic potentials. Application to formamide, methanol, and formic acid* **J. Phys. Chem.** (1978), 82, 592.
- (42) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model* **J. Phys. Chem.** (1993), 97, 10269.

Chapter 5

The solution structure of double helical arabino-nucleic acids (ANA and 2'F-ANA): effect of arabinoses in duplex-hairpin interconversion

RESUMEN

En este capítulo se ha determinado la primera estructura tridimensional de un dúplex autocomplementario que combina segmentos continuos de nucleótidos de ANA y 2'F-ANA. La secuencia utilizada es el dodecamero de Dickerson-Drew, con los nucleótidos de 2'F-ANA en el extremo y los de ANA en la zona central del oligonucleótido. Se ha observado que esta secuencia adopta dos estructuras en disolución, de estabilidad muy similar pero distinta molecularidad. La estructura monomérica es una horquilla, en la cual los nucleótidos de 2'F-ANA están apareados formando el “*stem*” y los nucleótidos de ANA se encuentran desapareados en el “*loop*”. La estructura bimolecular es un dúplex con todos los nucleótidos apareados (2'F-ANA y ANA), formando una hélice tipo B con todos los nucleótidos en conformación *sur*. En el segmento ANA:ANA, se observan interacciones desestabilizantes entre el grupo hidroxilo y el H8 de la base que provocan cambios en el ángulo glicosídico e impiden una geometría óptima en este tipo de pares de bases. En los pares 2'F-ANA:2'F-ANA, por el contrario, los nucleótidos de 2'F-ANA se acomodan muy bien en la estructura tipo B y permiten la formación de interacciones estabilizantes 2'F..H8.

ABSTRACT

In this chapter we report the first structure of double helical arabinonucleic acid (ANA), the C2'-stereoisomer of RNA, and the 2'-fluoro-arabinonucleic acid analogue (2'-F-ANA). A chimeric dodecamer based on the Dickerson sequence, containing a contiguous central segment of arabinonucleotides, flanked by two 2'-deoxy-2'-fluoro-arabinonucleotide wings (2'-F-ANA) was studied. Our data shows that this chimeric oligonucleotide can adopt two different structures of comparable thermal stabilities. One structure is a monomeric hairpin in which the stem is formed by base-paired 2'-F-ANA nucleotides and the loop by un-paired ANA nucleotides. The second structure is a bimolecular duplex, with all the nucleotides (2'-F-ANA and ANA) forming Watson-Crick base-pairs. The duplex structure is canonical B-form, with all arabinoses adopting a pure C2'-*endo* conformation. In the ANA:ANA segment, steric interactions involving the 2'-OH substituent provoke slight changes in the glycosidic angles and therefore, in the ANA:ANA base-pair geometry. These distortions are not present in the 2'-F-ANA:2'-F-ANA regions of the duplex, where the 2'-OH substituent is replaced by a smaller fluorine atom. 2'-F-ANA nucleotides fit very well into the geometry of B-form duplex, allowing for favorable 2'-F..H8 interactions. This interaction has been observed before in single 2'-F-ANA nucleotides and shares some features of pseudohydrogen bonds observed in 2'-F-ANA:RNA hybrids (chapter 6).

5.1 INTRODUCTION

The increasing number of promising applications of modified nucleic acids ranges from new antisense/siRNA therapies to the design of DNA-based nano-devices. For this reason, understanding the structural effects of nucleic acids substitutions is a field of enormous interest. Among the many substitutions in the sugar moiety, arabino nucleic acid (ANA) and its 2'-fluorinated derivative (2'-F-ANA; Figure 1) are particularly interesting¹⁻³. The attractiveness of arabinose modified oligonucleotides was first based on their nuclease resistance and their ability to bind to target mRNA and elicit enzymatic degradation of target mRNA through both RNase H and the RNA-induced silencing complex (RISC)^{1,4-7}. More recently, arabinonucleotide derivatives have been used as enhanced *in vivo* cellular DNA labels³. In today's era of synthetic biology, arabinonucleic acids have become attractive systems for studying the evolution of functional biopolymers by *in vitro* selection^{8,9}. Recent work on ANA, 2'-F-ANA and other nucleic acid analogues has been inspired by the desire to construct genetic systems based on alternative chemical platforms¹⁰.

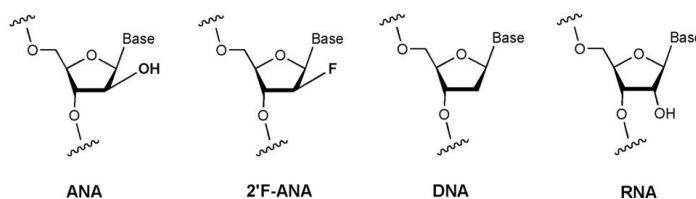


Figure 1: Structures of ANA and 2'-F-ANA in comparison with DNA and RNA

Despite the close similarity between ANA and 2'-F-ANA, their binding affinities for RNA are strikingly different^{1,6,11-13}. ANA binds to RNA with relatively low affinity, whereas 2'-F-ANA forms thermally stable hybrids with RNA. Besides its high binding affinity towards RNA, it has also been reported that 2'-F-ANA binds to itself to form B-like 2'-F-ANA:2'-F-ANA duplexes of high thermal stability^{14,15}. The structure of pure 2'-F-ANA and pure ANA duplexes remains largely unexplored. We show here that ANA, unlike its C2'-epimer (RNA) or its 2'-F cousin (2'-F-ANA), forms ANA:ANA duplexes of poor thermal stability. However, the appropriate

combination of 2′F-ANA and ANA nucleotides renders sufficient stability to an internal ANA:ANA segment to permit its characterization by NMR for the first time. We also report on the first structural characterization of a 2′F-ANA:2′F-ANA duplex. Specifically, we focused on determination of the three-dimensional structure of the chimeric dodecamer, 5′-fCfGfCfGaAaAaUaUfCfGfCfG-3′, through applying two-dimensional NMR spectroscopy experiments and restrained molecular dynamics.

5.2. RESULTS

5.2.1. Effect of ANA and 2′F-ANA modifications on duplex stability

Duplex formation and melting was monitored by UV spectroscopy. Besides the self-complementary sequences containing ANA modifications, isosequential RNA:RNA (**RR**) and DNA:DNA (**DD**) and 2′F-ANA:2′F-ANA (**FF**) control duplexes were also included in the UV melting experiments. T_m values are shown in Table 1.

Table 1: T_m values for the self-complementary chimeric duplexes (buffer conditions: 140 mM KCl, 1 mM MgCl₂, and 5 mM Na₂HPO₄, pH 7.2).

Name	Sequence (5′-3′) ^a	T_m (°C)
DD	5′-CGCGAATTCGCG-3′	58
RR	5′-CGCGAATTCGCG-3′	65
FF	5′-CGCGAATTCGCG-3′	76
AA	5′-CGCGAAUUCGCG-3′	-
alt(AF)	5′-CGCGAAUICGCG-3′	36
alt(DF)	5′-CGCGAATTCGCG-3′	60
gap(FA)	5′-CGCGAAUUCGCG-3′	56
gap(FD)	5′-CGCGAATTCGCG-3′	58

^aLegend: 2′F-ANA, RNA, DNA and ANA.

Whereas **FF** duplex, containing only 2′F-ANA nucleotides, exhibits the highest thermal stability, **AA** with only ANA nucleotides has a marginal stability and does not show any hypochromicity. Comparison between 1–1 altimer design, **alt(AF)**, or gapmer design, **gap(FA)**, with their corresponding control duplex **FF** also indicates that insertion of ANA nucleotides in the sequence is destabilizing. Oligonucleotide **gap(FA)** is especially interesting because while it contains a tract of destabilizing

ANA modifications, insertion of 2'-F-ANA nucleotides on flanks compensates this destabilization to the extent that permits the characterization of ANA:ANA base pairs by NMR techniques.

5.2.2. Duplex-hairpin equilibrium

NMR spectra of **gap(FA)** recorded at different oligonucleotide concentrations are indicative of self-complementary equilibrium. The imino region, shown in Figure 2, is particularly informative. Signals corresponding to AU base pairs are not observed at low concentration, suggesting the formation of a hairpin structure in which the stems contain 2'-F-ANA CG pairs, and the loop consists of four unpaired ANA nucleotides. At high oligonucleotide concentrations, however, the imino region is consistent with formation of a double stranded duplex structure in which all nucleotides (ANA and 2'-F-ANA) are base paired.

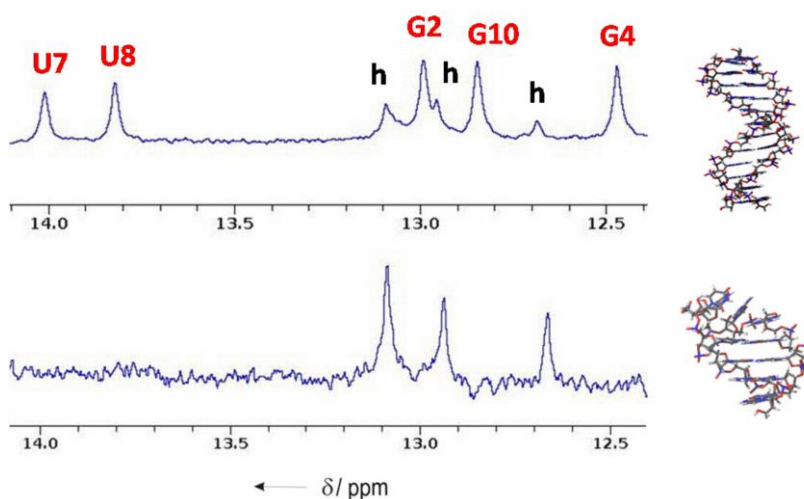


Figure 2: Imino region of the ^1H NMR spectra of 5'-fCfGfCfGaAaAaUaUfCfGfCfG-3' at two different oligonucleotide concentrations: Top: 0.9 mM and Bottom: 0.2 mM (buffer conditions: 100 mM NaCl, 25 mM sodium phosphate, pH = 7). Uppercase labels stands for duplex and lowercase labels for hairpin.

The observation of a hairpin species at low NMR concentrations suggests that the melting temperature of **gap(FA)** measured by UV at low oligonucleotide concentration corresponds, in fact, to hairpin denaturation and not to duplex dissociation as in the other oligonucleotides shown in Table 1 (the slightly different buffer conditions in UV

and NMR experiments do not affect the NMR spectra, data not shown). Fortunately, hairpin and duplex melting processes can be studied independently by ^1H or ^{19}F NMR, since the equilibrium between duplex and hairpin species exhibits a slow kinetics on the NMR time scale (Figure 3). ^{19}F NMR is particularly useful to monitor secondary structure transitions in nucleic acids^{16,17}. Both exchangeable ^1H and ^{19}F NMR spectra, recorded at different temperatures, indicate that these two species have similar melting temperatures at 0.8 mM oligonucleotide concentration.

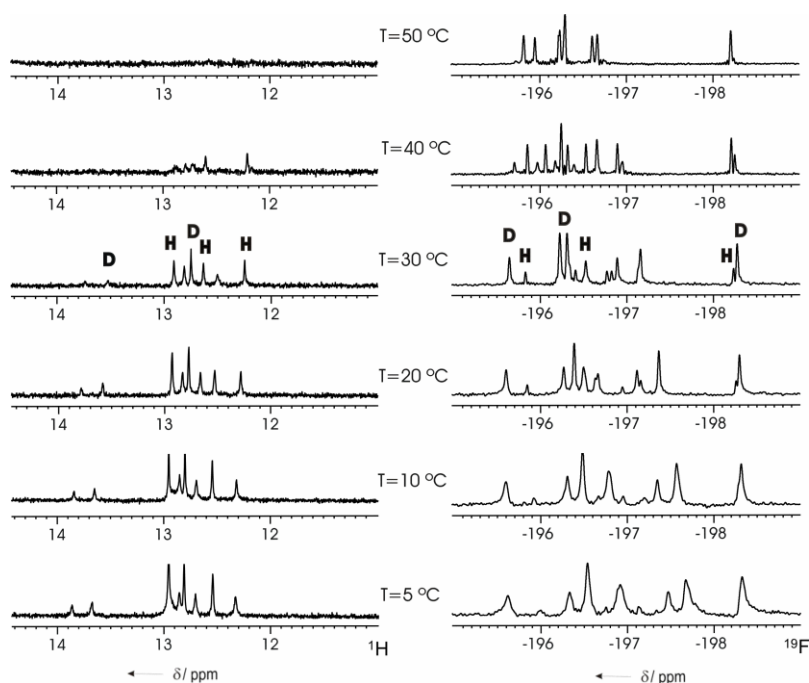


Figure 3: Imino region of the ^1H (left) and complete ^{19}F (right) NMR spectra of 5'-fCfGfCfGaAaAaUaUfCfGfCfG-3' (**gap(FA)**) at different temperatures. Signals of duplex (**D**) and hairpin (**H**) species are indicated (0.8 mM oligonucleotide concentration, in the same buffer conditions as in Figure 2).

The thermodynamic parameters for the hairpin to duplex equilibrium were estimated from van't Hoff analysis of the equilibrium constants that were obtained from the relative intensities of the ^{19}F NMR signals between equivalent protons in both forms (Figure 4). This analysis was carried out at temperatures somewhat below the T_m of both species, where the two-state approximation is valid. The resulting values for

$\Delta G_{298}^{H \rightarrow D}$, ΔH and $T\Delta S$ are -0.7 kcal/mol, -6.6 kcal/mol and -5.9 kcal/mol, respectively. This ΔG value indicates that the duplex is slightly more stable than the hairpin under these experimental conditions. ΔS is negative as expected, but its value is surprisingly small, which could be attributed to better hydration of the hairpin or a less organized duplex structure. The small negative enthalpy change suggests that contribution of ANA:ANA base-pairs to the stability of the duplex is only marginal.

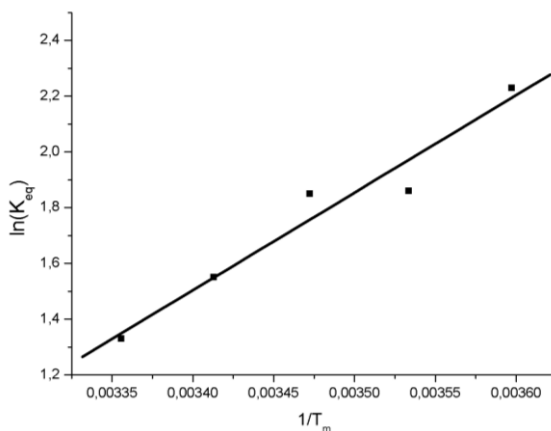


Figure 4: Van't Hoff plot of $\ln(K_{eq})$ vs $1/T$, used to derived thermodynamic parameters.

It is interesting to compare the NMR melting experiments of **gap(FA)** with that of the unmodified DNA control. As shown in Figure 5, all the imino proton signals of the DNA control (**DD**) disappear together at the same temperature, with the only exception for terminal base-pairs. For the **gap(FA)** sequence, however, ANA imino protons located in the central region of the duplex disappear more readily and at much lower temperatures compared to 2'-F-ANA imino protons. Such a behavior indicates that ANA:ANA base-pairs are less stable than DNA:DNA base pairs.

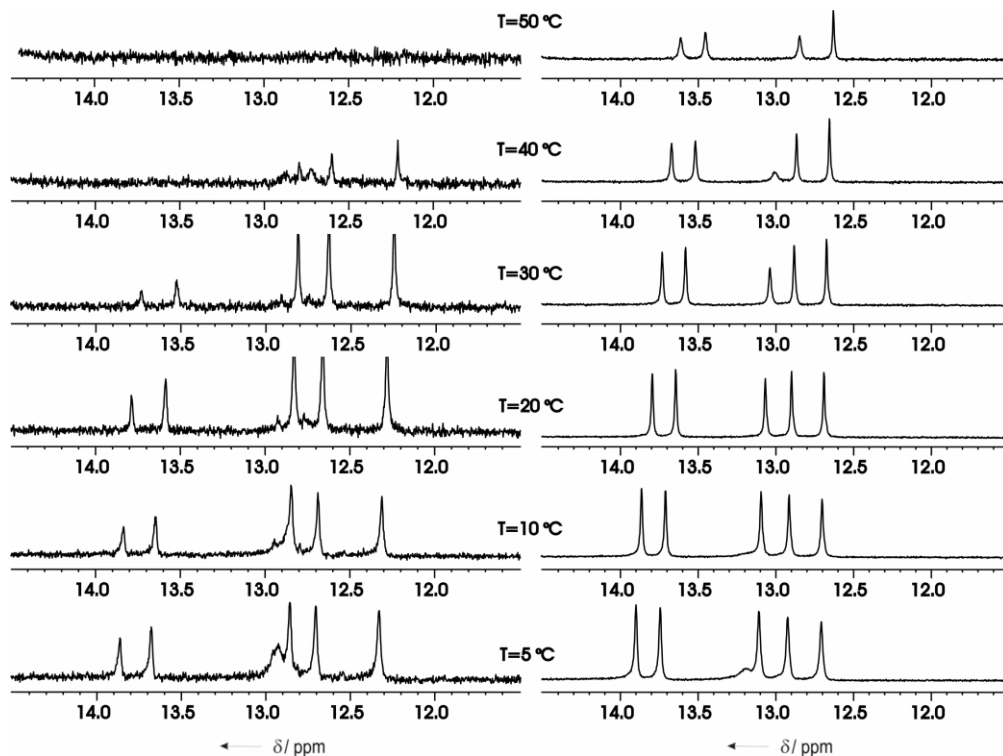


Figure 5: Left) ^1H spectra of 5'-fCfGfCfGaAaAaUaUfCfGfCfG-3' at different temperatures. Right) ^1H spectra of DNA sequence 5'-CGCGAATTCGCG-3' at different temperatures. Oligonucleotide concentration 0.8 mM, same buffer conditions as in Figure 2.

5.2.3. NMR assignments

Sequential assignments of exchangeable and non-exchangeable proton resonances were performed following standard methods for right-handed, double-stranded nucleic acids, using DQF-COSY, TOCSY and 2D NOESY spectra. At high oligonucleotide concentrations, the assignment pathways could be followed in the base-H1', and in the base-H2'' regions (Figure 6 and Table A5.1). Assignment of the duplex was partially hindered by the low intensity of the intra-nucleotide signals base-H1' and base-H2'', caused by the splitting of H2'' and H1' signals by fluorine. The assignment pathways of the low-concentration species cannot be followed between residues A6 to C9, due to the loss of sequential connectivities. In spite of these problems, the complete assignment of ^1H and ^{19}F resonances of both

species could be carried out by the analysis of experiments recorded at different temperatures and oligonucleotide concentrations. Except for some guanine and adenine amino resonances that could not be observed, most of the exchangeable protons of the duplex could be identified in the NOESY spectra in H₂O (Figure 7A). The observed cross-peak patterns indicate that all bases in the double-stranded duplex structure are forming Watson-Crick base-pairs. Only exchangeable protons corresponding to CG base pairs were observed in the low concentration hairpin species.

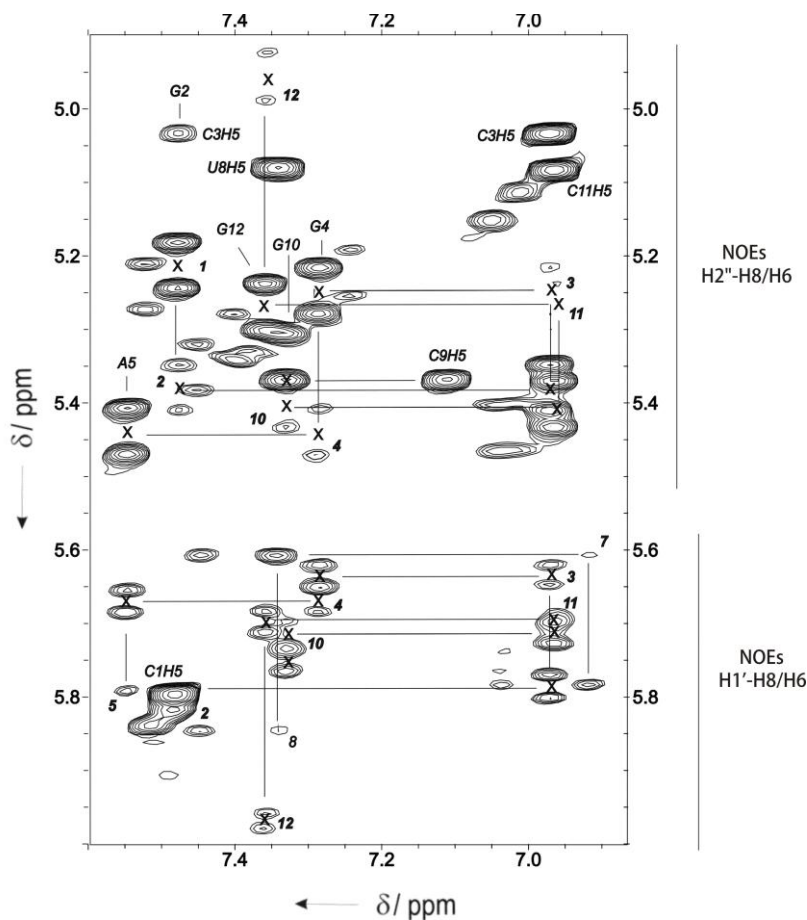


Figure 6: Region of the NOESY spectrum in D₂O (250 ms mixing time). H1'-base and H2''-base assignment pathways are indicated. Split signals are indicated with a cross in the middle of both peaks. Oligonucleotide concentration 0.8 mM, same buffer conditions as in Figure 2.

Assignment of ^{19}F resonances was carried out through their heteronuclear correlations with the adjacent $\text{H2}''$, $\text{H3}'$, $\text{H4}'$ and $\text{H1}'$ protons. Sequential and intra-residual ^{19}F - ^1H /H6/H8 cross-peaks were observed in the HOESY spectrum (Figure 7B).

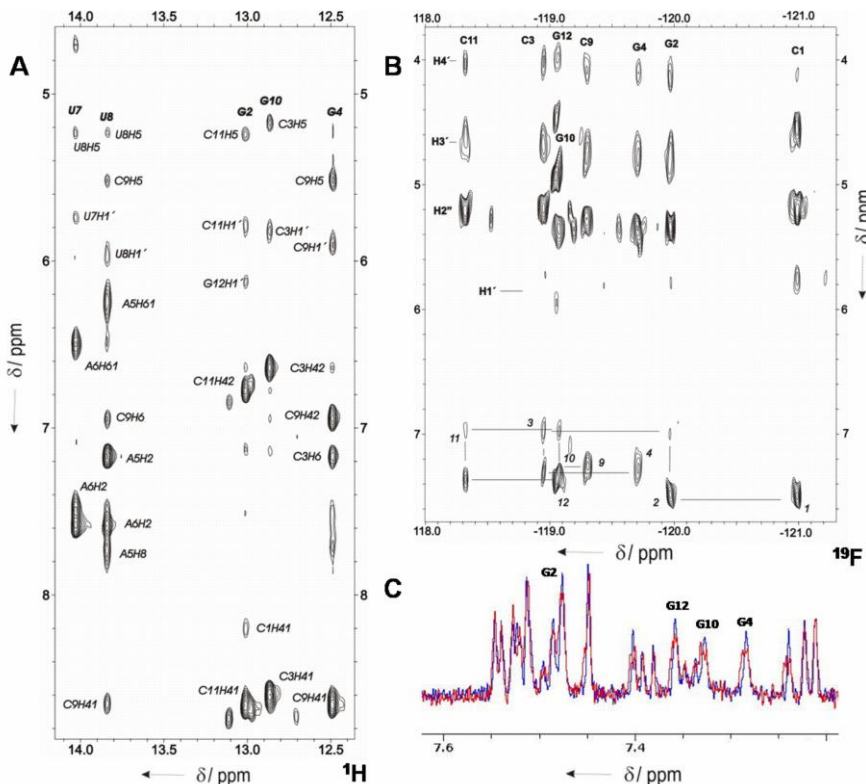


Figure 7: A) Region of the NOESY spectrum of *gap(FA)* in H_2O (150 ms mixing time). Watson-Crick base-pairs can be established for all the residues. B) HOESY spectrum in D_2O , showing the ^{19}F - ^1H assignment pathway. C) ^{19}F -coupled (red) and ^{19}F -decoupled (blue) ^1H NMR spectra, demonstrating different signal intensity in H8 2'- F -ANA purines. H8 of 2'- F -ANA purines in the duplex are labeled. Oligonucleotide concentration 0.8 mM, same buffer conditions as in Figure 2.

5.2.4. Experimental constraints and structure calculations

Quantitative distance constraints were obtained from NOESY experiments by using a complete relaxation matrix analysis. A total of 194 structurally relevant experimental distance constraints were obtained (see Table 2). The strong sequential $\text{H2}''$ -H6/8 NOEs confirm a B-like conformation. In addition to the NOE-

derived information, semi-quantitative analysis of the J -coupling constants obtained from DQF-COSY spectra was carried out. For 2'-F-ANA nucleotides, $J_{1'2''}$ are between 1 and 3 Hz, and $J_{3'2''}$ are almost zero, whereas $J_{1'2''}$ and $J_{3'2''}$ values are both negligible for ANA nucleotides (Table A5.2). $J_{F-H1'}$ and $J_{F-H3'}$ were estimated from the splitting in NOESY cross-peaks. Heteronuclear and homonuclear J -couplings are consistent with a South puckering in all arabinose nucleotides¹⁸. Accordingly, sugar torsion angles were constrained to avoid North conformations (allowing East and South). Backbone dihedral angles were not constrained.

Table 2: NMR restraints and structural statistics.

Experimental ^1H-^1H and ^{19}F-^1H distance constraints	
Total number	194
Intra-residue	83
Sequential	61
Inter-strand	32
^{19}F - ^1H	18
RMSD* (Å)	
all heavy atoms	0.9 ± 0.2
backbone	1.2 ± 0.3
all bases	0.8 ± 0.2
sum of violations (Å) (average and range)	3.5 (3.1..3.7)
max. violation (average and range)	0.6 (0.5..0.6)
NOE energy kcal/mol (average and range)	23.1 (21..26)

* All except terminal residues

In the case of 2'-F-ANA nucleotides, heteronuclear J -couplings were detected between 2'-fluorine and purine H8 of the same nucleotide (Figure 7C). These intra-residual couplings have been observed before in single 2'-F-ANA modified nucleotides and in hybrids with RNA (chapter 6)^{11,13,19}.

Distance and torsion angle constraints were used to calculate the structure by restrained molecular dynamics as described in Annex I. Twenty final duplex structures were obtained from an explicit solvent refinement and are displayed in Figure 8A. The calculations converge to a well-defined structure, with an RMSD of 0.9 Å (excluding the terminal residues). The final AMBER energies and NOE terms are reasonably low in all the structures, which do not exhibit significant constraint violations (see Table 2). Both 2'-F-ANA and ANA nucleotides are very well defined.

In the region of the 2'-F-ANA-ANA junction, nucleotides A5 and A17 exhibit larger conformational deviations.

5.2.5. Description of the duplex structure

The duplex structure of 5'-fCfGfCfGaAaAaUaUfCfGfCfG-3', **gap(FA)**, belongs to the B-form family. The RMSD between the average structure of **gap(FA)** and the canonical A-form structure of the same sequence is larger than the RMSD between **gap(FA)** and a canonical B-form (5.4 Å and 1.8 Å, respectively). In addition, helical parameters are characteristics of B-form helices, with an average rise of 3.3 Å and twist angles around 34°. A summary of helical parameters is shown in Table A5.3.

The geometry of arabinoses and 2'-F-arabinoses is very similar; most sugar pseudorotation phase angles range from -120° to -170° and glycosidic angles from -110° to -135° (excluding terminal residues) (Figure 8C,D). Backbone torsion angles are well defined and within the usual range for double helical structures. The C2'-*endo* sugar pucker of arabinose residues is in agreement with previous solution studies on DNA duplex with 2'-F-ANA substitutions²⁰, and contrasts with early crystallographic studies in which 2'-F-ANA adopt an O4'-*endo* conformation²¹. An interesting observation was the splitting of the H8 proton signals due to heteronuclear ¹⁹F-¹H J-couplings (Figure 7C, Table A5.2). These couplings likely reflect pseudohydrogen bonds^{13,19} as the intra-residual distances between the 2'-F and H8 protons are very short (<2.7 Å). Interestingly, intra-residual distances between the arabinose 2'-OH group and H8 protons are also very short (<2.6 Å) and we hypothesize that in this case, the 2'-OH..H8 interaction is destabilizing due to the larger size of the hydroxyl group (Figure 8E,F). It seems that the unfavorable steric contact between 2'-OH and H8 out weights their favorable electrostatic interaction shifting the glycosidic angles of arabinoses towards high anti values (-135°) (Figure 8D)¹². Unfavorable van der Waals contacts also cause distortions in the ANA:ANA base pairs, as shown in some base:base helical parameters, such as opening and inclination (see Table SA5.4). These distortions in ANA glycosidic angles partially hinder the co-planarity of A:U base pairs.

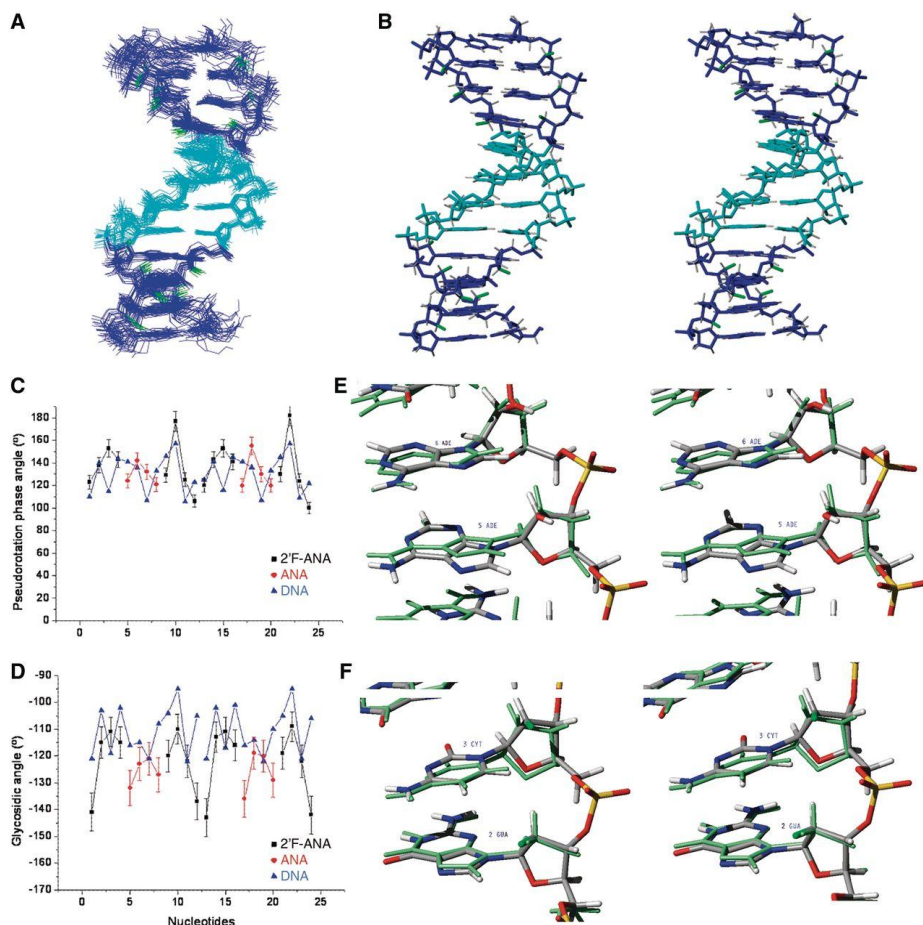


Figure 8: A) Ensemble of the 20 refined structures of the duplex form of 5'-fCfGfCfGaAaAaUaUfCfGfCfG-3'. B) Stereo view of the average structure. ANA nucleotides are shown in cyan, 2'F-ANA nucleotides in blue and fluorine atoms in green. C) Pseudorotation phase angle and D) glycosidic angle vs sequence. E) Stereo view showing details of ANA residues, and F) 2'F-ANA residues, superimposed to the unmodified DNA structure (in green color, pdb: 1DUF).

5.2.6. Hairpin structure

The hairpin structure is predominant only at low oligonucleotide concentrations, where conditions for obtaining structural constraints from NMR data are not optimal. Although most of the signals could be assigned, the number of distance constraints derived from NOESY experiments was low, impeding the determination of a high resolution structure. In spite of this limitation, some features of this

interesting unimolecular hairpin structure were deduced from the NMR data. The sequential sugar-base NOE connectivity could be traced for all 2'-F-ANA residues and A5 residue in the 2'-F-ANA-ANA interface, but not for other ANA residues. These data are consistent with a stem-loop hairpin structure composed of a four CG base-pair stem, formed by 2'-F-ANA nucleotides and a four-residue loop, formed by ANA nucleotides. Furthermore, strong sequential H2''-H6/8 NOEs are indicative of a B-like conformation for the 2'-F-ANA stem, which is very similar to the 2'-F-ANA:2'-F-ANA segment of the duplex structure. In case of ANA nucleotides, the lack of NOEs involving aromatic protons suggests that the bases are somewhat disordered. This is also supported by the chemical shifts of ANA H6/H8 protons, which exhibit very similar values as those of the fully denatured oligonucleotide observed at high temperatures. However, qualitative analysis of the J -coupling constants obtained from DQF-COSY spectra shows that $J_{1'2'}$ and $J_{2'3'}$ are small, with very similar values as those found in 2'-F-ANA and ANA nucleotides in the duplex (Figure 9). This observation indicates that arabino sugars have a similar rigidity in a duplex as in single-stranded loop, retaining a South conformation in both cases.

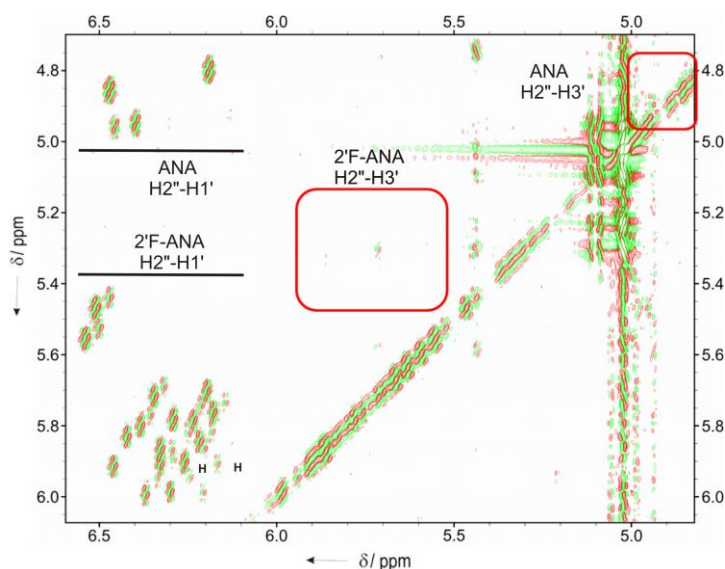


Figure 9: Region of the DQF-COSY spectrum in D_2O . $H1''$ - $H2''$ signals for 2'-F-ANA and ANA are at 5.8 and 4.9 ppm approximately. Weak $H2''$ - $H3''$ signals for 2'-F-ANA and ANA are indicated by red squares. Some of the $H2''$ - $H1''$ signals for the hairpin are labeled (H).

The structural features described above were confirmed by the model calculation shown in Figure 10. Unrestrained molecular dynamic trajectories show that the sugar-phosphate backbone of loop and stem regions of the hairpin is well defined, whereas ANA nucleobases in the loop are disordered.

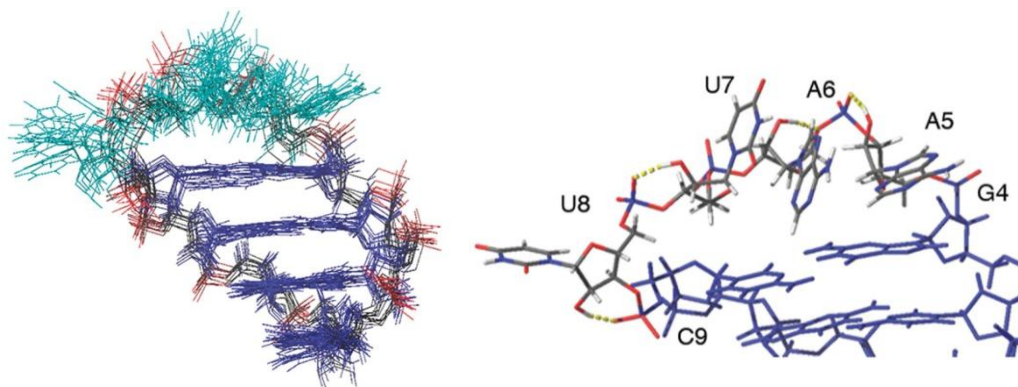


Figure 10: Model of the hairpin structure of **gap(FA)**. Left: Ensemble of 10 structures from molecular dynamics calculations. Right: Detail of the ANA residues in the loop, showing hydrogen bonds between the 2'-OH and phosphate oxygens.

5.3. DISCUSSION

5.3.1. Comparison of ANA and 2'-F-ANA with unmodified DNA structure

The Dickerson's dodecamer is probably the most studied DNA duplex in the literature²²⁻²⁵. Among the different structures deposited in the PDB, the structure obtained by Tjandra *et al.* (PDB: 1DUF) is the most appropriate model to use for comparison since it was also determined by NMR methods and with a very high resolution²⁴. Overall Tjandra's structure is very similar to the duplex conformation of **gap(FA)**. However, some differences are observed. The RMSD between average structure of **gap(FA)** duplex and Tjandra's average structure is 1.3 Å for heavy atoms and 0.9 Å for the bases. RMSD between ANA bases is 0.5 Å and between 2'-F-ANA bases is approximately 0.6 Å. The pseudorotation phase angles of ANA and 2'-F-ANA sugars are more South relative to the DNA sugars (Figure 8C). These results suggest that arabinose sugars are more rigid compared to deoxyribose sugars, which are in dynamic equilibrium between North and South conformations. In fact, observed $J_{1'2''}$ coupling values (Table A5.2) convinced us

that such North-South conformational transitions do not occur in ANA and 2′F-ANA nucleotides, as even a very small population of North conformation would give rise to a larger value for $J_{1'2'}$ couplings. Our findings are consistent with NMR studies on ANA:RNA (PDB: 2KP3) and 2′F-ANA:RNA (PDB: 2KP4) hybrids, which also show a quite rigid arabinose conformations (chapter 6)¹³.

Differences in glycosidic angles between **gap(FA)** and Tjandras’s structures are limited to the ANA residues mentioned above. Backbone dihedral angles are also similar (Table A5.5, A5.6). Exceptions are the zeta angles, which present smaller values than in the solution structures of unmodified DNA Dickerson’s dodecamer²⁴. This is caused by the repulsion between the 2′-substituent and the adjacent phosphate groups. Thus, sequential P-H2′ distances in Tjandra’s structure are smaller than 4 Å, whereas in **gap(FA)** these distances (sequential P-2′F and P-2′O) are around 5 Å. The minor groove width is narrower than in canonical B-form helices in solution (Figure 11). Interestingly, a minor groove narrowing is observed in the central part of the sequence. A similar effect was described for the AATT moiety in the unmodified DNA helix²⁴.

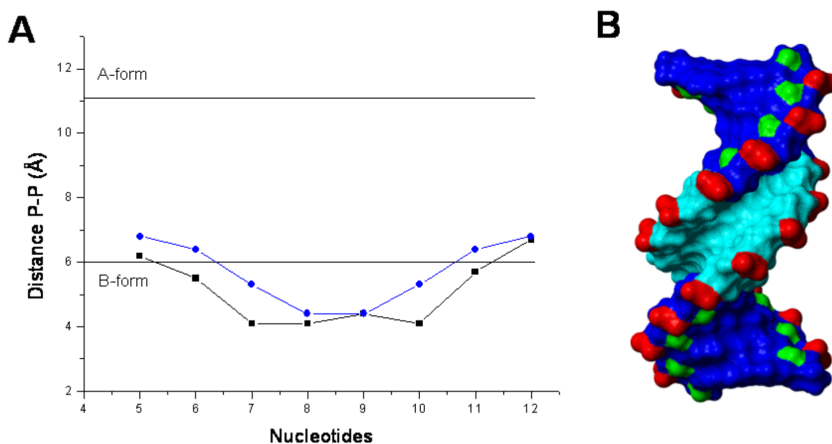


Figure 11: A) Minor groove width vs sequence and B) accessible surface showing the minor groove narrowing observed in the central part of the molecule. **Gap(FA)** structure parameters are shown in black and unmodified DNA values (pdb: 1DUF). Surface corresponding to ANA residues is shown in cyan, 2′F-ANA residues in blue, phosphates in red and fluorine atoms in green.

5.3.2. ANA stabilizes stem-loop structures

All palindromic sequences can, in principle, form stem-loop hairpin structures, although such structures are mainly formed at much lower salt and oligonucleotide concentrations than those used in this study. Very early spectroscopic studies on Dickerson's sequence suggested the formation of alternative conformations^{26,27}. However, they only occur at much lower concentrations than those required in structural studies. In high resolution NMR and crystallographic studies unmodified DNA Dickerson's dodecamer has been found as a duplex. In addition, most DNA analogs studied with this sequence adopt a predominant duplex structure. An exception is the recently reported crystallographic structure of Dickerson's dodecamer containing N-methanocarpa-thymineDNA analogs²⁸. However, this modification does not induce a detectable hairpin formation in solution^{29,30}. The pure RNA version of Dickerson's sequence also adopt a duplex structure in solution³¹. However, several modifications on the RNA sequence have been reported to induce hairpin formation³².

In the case of ANA modification, the unfavorable steric interactions due to short –OH to aromatic proton distances observed in the duplex are partially alleviated in the loop structure. Such a reduction of unfavorable steric clashes has an effect in the energetic balance between the duplex and the hairpin forms, and is partially responsible for the similar stability of the two forms. Furthermore, several intra-residual hydrogen bonds between 2'-OH and phosphate oxygens were observed for the arabinose loop residues during the unrestrained molecular dynamic calculations (Figure 10). Intra O2'-H2'...O5' hydrogen bonds have previously been observed in ANA nucleosides^{12,33} and stabilize the South sugar conformation. It must be noticed that these interactions are not possible in a B-form helical conformation.

Furthermore, arabinose 2'-OH groups are more solvent exposed in the hairpin-loop than in the duplex species. Therefore dehydration of these groups is required for duplex formation, but not for hairpin folding. This effect may be responsible for the small entropy difference between the two conformations. Solvation of 2'-OH groups is an important contribution to the stability of RNA hairpins. Stem-loop

hairpins are more common in RNA than in DNA³⁴. It is not uncommon that short oligoribonucleotides adopt duplex or hairpin structures of comparable stabilities and transition between them are related to different biological processes. This observation raises the question whether ANA is a DNA or an RNA mimic. In the context of double helical structures, ANA is clearly a DNA analog (i.e. C2'-*endo* sugar conformations, B-type duplex). However, ANA promotes formation of stem-loop structures and, in this sense, behaves as an RNA analog. This tendency of ANA nucleotide tracts to form loop structures may be useful for the rational design of nucleic acid hairpins or in applications where a duplex/hairpin conformational switch is desirable.

5.3.3. ANA vs 2'-F-ANA duplexes – why such a difference in stability?

ANA and 2'-F-ANA segments of **gap(FA)** duplex exhibit very similar structures. Sugar and glycosidic angle conformations provoke a systematic steric clash between the 2'-OH and their own H6/H8 base protons in ANA nucleotides. However, the same geometry provokes favorable electrostatic interactions with 2'-F-ANA nucleotides, due to the smaller radii of the fluorine and its higher electronegativity, compared to the 2'-OH group. Similar effects have been observed for hybrids of 2'-F-ANA and ANA with RNA¹³, as will be described in chapter 6. In 2'-F-ANA:RNA hybrids, the geometry of the C-2'F/H8 network is almost co-linear, facilitating the formation of sequential pseudo-hydrogen bonds. The 2'F..H8 contacts are particularly favorable in pyrimidine-purine steps where the base stacking geometry can adjust to optimize the interaction without incurring a steric penalty. In the duplex structure of **gap(FA)**, 2'F..H8 interactions are intra-residual although the geometry is not so favorable (the 2'F..H6/8-C angle is around 100-110°).

Organic fluorine is considered to be a poor hydrogen bond acceptor³⁵⁻⁴¹, however there is evidence that fluorine-mediated pseudohydrogen bonding is important in duplex thermal stability^{13,42-44}. Since in **gap(FA)** these interactions are intra-residual, they can occur both in the duplex and in the single stranded denatured states and, consequently, their contribution to duplex stability is not direct. Most probably, such favorable non-covalent interactions act by preorganizing the bases

and therefore reducing the entropic penalty of duplex formation. Alternatively, duplex formation may facilitate intra-residual pseudohydrogen bonding, thus making duplexation enthalpically favorable. Finally, it has been observed that 2'-substituents in riboses modulate base-pairing strength and π - π stacking interactions in 2'-F-RNA structures⁴⁵. Fluorine substitution in arabinoses may also polarize their nucleobases and affect the strength of Watson-Crick hydrogen bonds in 2'-F-ANA:2'-F-ANA base pairs.

5.3.4. Alternative genetic systems based on arabinose?

Recent work on nucleic acid analogues have been inspired by the desire to construct genetic systems based on alternative chemical platforms. Holliger and co-workers¹⁰ and Prof. Damha's laboratories^{8,9} have reported that DNA polymerases can catalyse DNA synthesis on a 2'-F-ANA template strand. Conversely, DNA polymerases are able to synthesize ANA, 2'-F-ANA or chimeric 2'-F-ANA-DNA strands on template DNA strands. Since ANA and 2'-F-ANA polymers can be copied to and from DNA with good fidelity, this attribute is sufficient to carry out the directed evolution of functional arabinonucleic acid molecules.

It is generally assumed that stable duplex formation is a crucial aspect of directed evolution of functional biopolymers based on a synthetic nucleic acid system. Although the poor stability of ANA duplexes would suggest that ANA may not be used to store or propagate information, there is ample evidence that polymerases are able to catalyze oligonucleotide synthesis in the absence of a stable product-template duplex⁴⁶. Consistent with this observation, an engineered D4K polymerase is able to carry out template-dependent DNA synthesis on an ANA template¹⁰, despite the poor stability of ANA:DNA duplexes¹. 2'-F-ANA has been copied to its own complement¹⁰ but this process is less efficient than copying information between 2'-F-ANA to and from DNA⁸⁻¹⁰, pointing to subtle structural differences between DNA and 2'-F-ANA duplexes as revealed in this study.

5.4. CONCLUSION

Arabinonucleic acids can adopt duplex structures in solution very similar to standard B-form DNA duplexes. Arabinoses and their fluorine-substituted analog, 2'-fluoroarabinose adopt pure C2'-*endo* conformations, with very little variation along the sequence. In spite of these structural similarities, the thermal stability of ANA and 2'-F-ANA duplexes is very different. Pure 2'-F-ANA duplexes are substantially more stable than DNA or RNA duplexes of the same sequence, whereas ANA:ANA duplexes are weak at best (duplex thermal stability increasing from ANA < DNA < RNA < 2'-F-ANA).

The electronegativity of fluorine provokes the polarization of the C-H8 bonds favoring the formation of intra-residue C-2'F...H8-C interactions. The chimeric oligonucleotide sequence, **gap(FA)**, adopts a duplex and a hairpin structure of surprisingly similar thermal stability. The co-existence of these two species, as observed independently by NMR methods, offered a good opportunity to study the structure of the ANA:ANA duplex. Analysis of the ANA residues in the loop or duplex structure indicate they are rigid and preorganized in a conformation that cannot adjust to optimize the base-stacking and Watson-Crick hydrogen bond formation without incurring a steric penalty. NMR studies of other single stranded arabino and fluoro-arabino oligonucleotides would help test and refine this hypothesis.

BIBLIOGRAPHY

- (1) Noronha, A. M.; Wilds, C. J.; Lok, C.-N.; Viazovkina, K.; Arion, D.; Parniak, M. A.; Damha, M. J. *Synthesis and Biophysical Properties of Arabinonucleic Acids (ANA): Circular Dichroic Spectra, Melting Temperatures, and Ribonuclease H Susceptibility of ANA-RNA Hybrid Duplexes* **Biochemistry** (2000), 39, 7050.
- (2) Watts, J. K.; Damha, M. J. 2'-F-Arabinonucleic acids (2'-F-ANA) — History, properties, and new frontiers **Can. J. Chem.** (2008), 86, 641.
- (3) Neef, A. B.; Luedtke, N. W. *Dynamic metabolic labeling of DNA in vivo with arabinosyl nucleosides* **Proc. Natl. Acad. Sci. U. S. A.** (2011), 108, 20404.
- (4) Dowler, T.; Bergeron, D.; Tedeschi, A.-L.; Paquet, L.; Ferrari, N.; Damha, M. J. *Improvements in siRNA properties mediated by 2'-deoxy-2'-fluoro-β-D-arabinonucleic acid (FANA)* **Nucleic Acids Res.**, 34, 1669.
- (5) Kalota, A.; Karabon, L.; Swider, C. R.; Viazovkina, E.; Elzagheid, M.; Damha, M. J.; Gewirtz, A. M. 2'-deoxy-2'-fluoro-β-D-arabinonucleic acid (2'-F-ANA) modified oligonucleotides (ON) effect highly efficient, and persistent, gene silencing **Nucleic Acids Res** (2006), 34, 451.

- (6) Li, F.; Sarkhel, S.; Wilds, C. J.; Wawrzak, Z.; Prakash, T. P.; Manoharan, M.; Egli, M. *2'-Fluoroarabino- and Arabinonucleic Acid Show Different Conformations, Resulting in Deviating RNA Affinities and Processing of Their Heteroduplexes with RNA by RNase H* **Biochemistry** (2006), 45, 4141.
- (7) Denisov, A. Y.; Noronha, A. M.; Wilds, C. J.; Trempe, J.-F.; Pon, R. T.; Gehring, K.; Damha, M. J. *Solution structure of an arabinonucleic acid (ANA)/RNA duplex in a chimeric hairpin: comparison with 2'-fluoro-ANA/RNA and DNA/RNA hybrids* **Nucleic Acids Res.** (2001), 29, 4284.
- (8) Peng, C. G.; Damha, M. J. *Probing DNA polymerase activity with stereoisomeric 2'-fluoro-β-D-arabinose (2'-araNTPs) and 2'-fluoro-β-D-ribose (2'-rNTPs) nucleoside 5'-triphosphates* **Can. J. Chem.** (2008), 86, 881.
- (9) Peng, C. G.; Damha, M. J. *Polymerase-Directed Synthesis of 2'-Deoxy-2'-fluoro-β-D-arabinonucleic Acids* **J. Am. Chem. Soc.** (2007), 129, 5310.
- (10) Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P. *Synthetic Genetic Polymers Capable of Heredity and Evolution* **Science** (2012), 336, 341.
- (11) Wilds, C. J.; Damha, M. J. *2'-Deoxy-2'-fluoro-β-d-arabinonucleosides and oligonucleotides (2'-F-ANA): synthesis and physicochemical studies* **Nucleic Acids Res.** (2000), 28, 3625.
- (12) Gao, Y. G.; Van der Marel, G. A.; Van Boom, J. H.; Wang, A. H. J. *Molecular structure of a DNA decamer containing an anticancer nucleoside arabinosylcytosine: conformational perturbation by arabinosylcytosine in B-DNA* **Biochemistry** (1991), 30, 9922.
- (13) Watts, J. K.; Martín-Pintado, N.; Gómez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; González, C.; Damha, M. J. *Differential stability of 2'-F-ANA•RNA and ANA•RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility* **Nucleic Acids Res.** (2010), 38, 2498.
- (14) Deleavey, G. F.; Watts, J. K.; Alain, T.; Robert, F.; Kalota, A.; Aishwarya, V.; Pelletier, J.; Gewirtz, A. M.; Sonenberg, N.; Damha, M. J. *Synergistic effects between analogs of DNA and RNA improve the potency of siRNA-mediated gene silencing* **Nucleic Acids Res.** (2010), 38, 4547.
- (15) Wilds, C. J. *Ph.D. thesis* (2000).
- (16) Graber, D.; Moroder, H.; Micura, R. *19F NMR Spectroscopy for the Analysis of RNA Secondary Structure Populations* **J. Am. Chem. Soc.** (2008), 130, 17230.
- (17) Kreutz, C.; Kählig, H.; Konrat, R.; Micura, R. *Ribose 2'-F Labeling: A Simple Tool for the Characterization of RNA Secondary Structure Equilibria by 19F NMR Spectroscopy* **J. Am. Chem. Soc.** (2005), 127, 11558.
- (18) Thibaudeau, C.; Plavec, J.; Chattopadhyaya, J. *A New Generalized Karplus-Type Equation Relating Vicinal Proton-Fluorine Coupling Constants to H-C-C-F Torsion Angles* **J. Org. Chem.** (1998), 63, 4967.
- (19) Anzahae, M. Y.; Watts, J. K.; Alla, N. R.; Nicholson, A. W.; Damha, M. J. *Energetically Important C-H...F-C Pseudohydrogen Bonding in Water: Evidence and Application to Rational Design of Oligonucleotides with High Binding Affinity* **J. Am. Chem. Soc.** (2010), 133, 728.
- (20) Ikeda, H.; Fernandez, R.; Barchi, J. J.; Huang, X.; Marquez, V. E.; Wilk, A. *The effect of two antipodal fluorine-induced sugar puckers on the conformation and stability of the Dickerson-Drew dodecamer duplex [d(CGCGAATTCGCG)]₂* **Nucleic Acids Res.** (1998), 26, 2237.
- (21) Berger, I.; Tereshko, V.; Ikeda, H.; Marquez, V. E.; Egli, M. *Crystal structures of B-DNA with incorporated 2'-deoxy-2'-fluoro-arabino-furanosyl thymine: Implications of conformational preorganization for duplex stability* **Nucleic Acids Res.** (1998), 26, 2473.
- (22) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Structure of a B-DNA dodecamer: conformation and dynamics* **Proc. Natl. Acad. Sci. U. S. A.** (1981), 78, 2179.
- (23) Tereshko, V.; Minasov, G.; Egli, M. *The Dickerson-Drew B-DNA Dodecamer Revisited at Atomic Resolution* **J. Am. Chem. Soc.** (1998), 121, 470.
- (24) Tjandra, N.; Tate, S.-i.; Ono, A.; Kainosho, M.; Bax, A. *The NMR Structure of a DNA Dodecamer in an Aqueous Dilute Liquid Crystalline Phase* **J. Am. Chem. Soc.** (2000), 122, 6190.
- (25) Kuszewski, J.; Schwieters, C.; Clore, G. M. *Improving the Accuracy of NMR Structures of DNA by Means of a Database Potential of Mean Force Describing Base-Base Positional Interactions* **J. Am. Chem. Soc.** (2001), 123, 3903.
- (26) Marky, L. A.; Blumenfeld, K. S.; Kozlowski, S.; Breslauer, K. J. *Salt-dependent conformational transitions in the self-complementary deoxydodecanucleotide d(CGCAATTCGCG): Evidence for hairpin formation* **Biopolymers** (1983), 22, 1247.
- (27) Miller, M.; Kirchhoff, W.; Schwarz, F.; Appella, E.; Chiu, Y.-y. H.; Cohen, J. S.; Sussman, J. L. *Conformational transitions of synthetic DNA sequences with inserted bases, related to the dodecamer d(CGCGAATTCGCG)* **Nucleic Acids Res.** (1987), 15, 3877.
- (28) Pallan, P. S.; Marquez, V. E.; Egli, M. *The Conformationally Constrained N-Methanocarba-dT Analogue Adopts an Unexpected C4'-exo Sugar Pucker in the Structure of a DNA Hairpin* **Biochemistry** (2012), 51, 2639.

- (29) Wu, Z.; Maderia, M.; Barchi, J. J.; Marquez, V. E.; Bax, A. *Changes in DNA bending induced by restricting nucleotide ring pucker studied by weak alignment NMR spectroscopy* **Proc. Natl. Acad. Sci. U. S. A.** (2005), 102, 24.
- (30) Maderia, M.; Shenoy, S.; Van, Q. N.; Marquez, V. E.; Barchi, J. J. *Biophysical studies of DNA modified with conformationally constrained nucleotides: comparison of 2'-exo (north) and 3'-exo (south) 'locked' templates* **Nucleic Acids Res.** (2007), 35, 1978.
- (31) Chou, S. H.; Flynn, P.; Reid, B. *Solid-phase synthesis and high-resolution NMR studies of two synthetic double-helical RNA dodecamers: r(CGCGAAUUCGCG) and r(CGCGUAUACGCG)* **Biochemistry** (1989), 28, 2422.
- (32) Plevnik, M.; Gdaniec, Z.; Plavec, J. *Solution structure of a modified 2',5'-linked RNA hairpin involved in an equilibrium with duplex* **Nucleic Acids Res.** (2005), 33, 1749.
- (33) Venkateswarlu, D.; Ferguson, D. M. *Effects of C2'-Substitution on Arabinonucleic Acid Structure and Conformation* **J. Am. Chem. Soc.** (1999), 121, 5609.
- (34) Varani, G. *Exceptionally Stable Nucleic Acid Hairpins* **Annu. Rev. Biophys. Biomol. Struct.** (1995), 24, 379.
- (35) Zhou, P.; Zou, J.; Tian, F.; Shang, Z. *Fluorine Bonding — How Does It Work In Protein-Ligand Interactions?* **J. Chem. Inf. Comput. Mod.** (2009), 49, 2344.
- (36) Dunitz, J. D. *Organic Fluorine: Odd Man Out* **ChemBioChem** (2004), 5, 614.
- (37) Böhm, H.-J.; Banner, D.; Bendels, S.; Kansy, M.; Kuhn, B.; Müller, K.; Obst-Sander, U.; Stahl, M. *Fluorine in Medicinal Chemistry* **ChemBioChem** (2004), 5, 637.
- (38) Ojala, W. H.; Skrypek, T. M.; MacQueen, B. C.; Ojala, C. R. *Intermolecular C-F...H-C contacts in the molecular packing of three isostructural N-(fluorophenyl)mannopyranosylamines* **Acta Crystallogr. Sect. C** (2010), 66, o565.
- (39) Carosati, E.; Sciabola, S.; Cruciani, G. *Hydrogen Bonding Interactions of Covalently Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field* **J. Med. Chem.** (2004), 47, 5114.
- (40) Mehta, G.; Sen, S. *Probing Fluorine Interactions in a Polyhydroxylated Environment: Conservation of a C-F...H-C Recognition Motif in Presence of O-H...O Hydrogen Bonds* **Eur. J. Org. Chem** (2010), 2010, 3387.
- (41) Mele, A.; Vergani, B.; Viani, F.; Meille, S. V.; Farina, A.; Bravo, P. *Experimental Evidence for Intramolecular Attractive Nonbonded C-F...H-C Interactions in 2',3'-Dideoxy-4'-(fluoromethyl)nucleosides — Through-Space JCF and JHF NMR Coupling Constants, Correlation with Empirical Parameters of Solvent Polarity and Single-Crystal X-ray Structures* **Eur. J. Org. Chem** (1999), 1999, 187.
- (42) Sun, Z.; McLaughlin, L. W. *Probing the Nature of Three-Centered Hydrogen Bonds in Minor-Groove Ligand-DNA Interactions: The Contribution of Fluorine Hydrogen Bonds to Complex Stability* **J. Am. Chem. Soc.** (2007), 129, 12531.
- (43) Pallan, P. S.; Egli, M. *Pairing Geometry of the Hydrophobic Thymine Analogue 2,4-Difluorotoluene in Duplex DNA as Analyzed by X-ray Crystallography* **J. Am. Chem. Soc.** (2009), 131, 12548.
- (44) Parsch, J.; Engels, J. W. *C-F...H-C Hydrogen Bonds in Ribonucleic Acids* **J. Am. Chem. Soc.** (2002), 124, 5664.
- (45) Patra, A.; Paolillo, M.; Charisse, K.; Manoharan, M.; Rozners, E.; Egli, M. *2'-Fluoro RNA Shows Increased Watson-Crick H-Bonding Strength and Stacking Relative to RNA: Evidence from NMR and Thermodynamic Data* **Angew. Chem. Int. Ed. Engl.** (2012), 51, 11863.
- (46) Tsai, C.-H.; Chen, J.; Szostak, J. W. *Enzymatic synthesis of DNA on glycerol nucleic acid templates without stable duplex formation between product and template* **Proc. Natl. Acad. Sci. U. S. A.** (2007), 104, 14598.

Chapter 6

The solution structure of 2'F-ANA:RNA and ANA:RNA hybrid duplexes. Roles of structure and hydrogen bonding in their different stability

RESUMEN

En este capítulo se describe la estructura tridimensional de dos dúplex híbridos 2'F-ANA:RNA y ANA:RNA. Ambos adoptan estructuras muy similares pero tienen estabilidades muy diferentes. Comparando la estructura de ambas secuencias hemos determinado que la estabilización de 2'F-ANA se debe principalmente a la formación de las interacciones 2'F..H8 y FC-H2''..O4, mientras que en el caso de ANA la desestabilización se debe a una serie de interacciones desfavorables entre el grupo hidroxilo y el H8 de la base secuencial. Tanto los nucleótidos de 2'F-ANA como los de ANA adoptan una conformación sudeste en los dúplex híbridos. Globalmente, la estructura de estos híbridos es intermedia entre A y B, siendo más parecida al tipo B en el caso de ANA. La cadena de 2'F-ANA, además, se encuentra preorganizada debido a la formación de pseudoenlaces de hidrógeno 2'F..H8, lo que también provoca un efecto estabilizante en la estructura.

ABSTRACT

Hybrids of RNA with arabinonucleic acids 2'F-ANA and ANA have very similar structures but strikingly different thermal stabilities. In this chapter we present our NMR studies on a fully modified 10-mer hybrid duplex. Comparison between the solution structure of 2'F-ANA:RNA and ANA:RNA hybrids indicates that the increased binding affinity of 2'F-ANA is related to several subtle differences, most importantly favorable (2'F..H8) and F-CH..O interactions which contrast with unfavorable internucleotide 2'-OH..base steric interactions in the case of ANA. While both 2'F-ANA and ANA strands maintained conformations in the southern/eastern sugar pucker range, the 2'F-ANA strand's structure was more compatible with the A-like structure of a hybrid duplex. Finally, while the two hybrid duplexes are of similar rigidities, 2'F-ANA single strands may be more suitably preorganized for duplex formation, due to the stabilizing interactions. Thus, we can conclude that the dramatically increased stability of 2'F-ANA:RNA and ANA:RNA duplexes is caused mainly by important differences in structure and hydrogen bonding.

6.1. INTRODUCTION

The arabinonucleic acids ANA and 2′F-ANA are very close cousins. Where ANA contains a hydroxyl group, 2′F-ANA contains fluorine (Figure 1A). The most extensive structural studies of ANA:RNA and 2′F-ANA:RNA duplexes to date, performed on short hairpins with modified stems, observed no major conformational differences between them, neither in sugar pucker nor hydrogen bonding nor steric effects^{1,2}. ANA and 2′F-ANA hybrid duplexes with RNA display structure and flexibility patterns that make them effective mimics of the DNA:RNA hybrid, even in terms of RNase H degradation³⁻⁵.

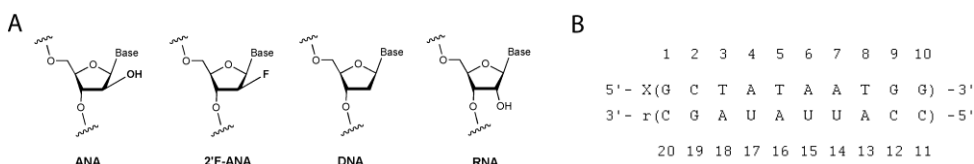


Figure 1: (A) Structures of ANA and 2′F-ANA in comparison with DNA and RNA. (B) Numbering scheme of the 10-mer hybrid duplexes used for this study. *X* indicates the DNA, 2′F-ANA or ANA strand. The corresponding hybrids are named **DR**, **FR** and **AR**, respectively. Thymines are replaced by uracils in the ANA strand⁶.

And so it is a puzzle that their binding affinities are strikingly different: ANA has relatively low affinity for RNA⁷, while 2′F-ANA binds to RNA with high affinity⁵. Because of this high binding affinity, and many other favorable properties, 2′F-ANA has shown promise for applications as diverse as gene silencing therapeutics, diagnostics and aptamer design⁸, and a 2′F-ANA-based antisense drug has received approval to begin clinical trials⁹. The elusive origin of the difference in binding affinity is therefore of significant interest: it will deepen our understanding of nucleic acid structure and binding while potentially allowing the design of derivatives with improved properties.

Initially, it was supposed that an unfavorable steric interaction involving the 2′-OH of ANA was responsible for its low binding affinity^{5,7}. Other studies have suggested that there could be a major difference in sugar pucker or hydrogen bonding between the two analogues^{10,11}. However, these structural explanations were

deemed unlikely based on the most significant structural studies of ANA:RNA and 2'F-ANA:RNA duplexes to date: high-resolution NMR studies that examined short hairpins containing a 4-bp hybrid stem and a 4-nt DNA loop^{1,2}. No major structural differences between ANA and 2'F-ANA were observed during this study, and the authors suggested differential hydration as a possible explanation for the greater stability of 2'F-ANA:RNA duplexes. However, we wondered if subtle differences might have been overlooked in the hairpin-based structural studies: while use of hairpins was convenient, the model contained only four hybrid base pairs, two of which might be affected by their proximity to the loop structure or the terminus. Therefore, to gain further insight on 2'F-ANA and ANA substitution in hybrid duplexes, we decided to carry out a structural and biophysical study of the decamer shown in Figure 1B. The structure of the unmodified control DNA:RNA duplex has been extensively studied by NMR and restrained molecular dynamics calculations, using conventional and time averaged constraints¹². This duplex is ideal for studying the origin of the difference in thermal stability between ANA:RNA and 2'F-ANA:RNA because it is representative of prototypical hybrids and because the difference in T_m values is nearly 18 °C (1.8 °C per base pair). Prof. Masad Damha's laboratory also verified that the duplex is a substrate of RNase H to ensure the highest applicability of the current study to the field of antisense therapeutics.

6.2. RESULTS

6.2.1. Duplex formation and thermodynamics

Duplex formation and melting was monitored by UV, CD and NMR spectroscopy. Besides the three hybrid duplexes used for NMR structural studies (Figure 1B), isosequential RNA:RNA (**RR**) and DNA:DNA (**DD**) controls were included in the melting experiments. T_m values are shown in Table 1. The 2'F-ANA:RNA (**FR**) duplex had the highest thermal stability of the five (T_m = 51.2 °C) while the ANA:RNA (**AR**) duplex was the least stable (T_m = 32.4 °C).

Table 1: T_m for the duplexes **DR**, **AR**, **FR**, **DD** and **RR**

Name	T_m (°C)
DR	33.7
AR	32.4
FR	51.2
DD	37.1
RR	46.9

NMR spectra are also consistent with duplex formation in all cases. Melting temperature experiments carried out on the NMR samples (by observing signals from imino protons upon increasing temperature) confirmed that the **FR** duplex is much more stable than **DR** or **AR** under the conditions used for NMR experiments (Figure 2).

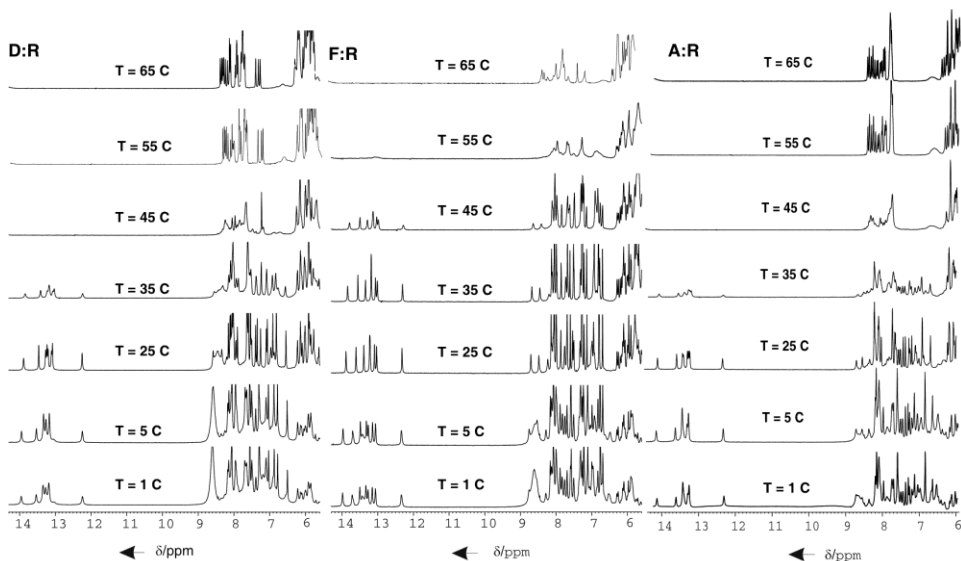


Figure 2: One dimensional NMR spectra of **FR**, **AR** and **DR** duplexes in H_2O at different T_s (0.8 mM oligonucleotide concentration, 25 mM sodium phosphate buffer, 100 mM NaCl, pH=7).

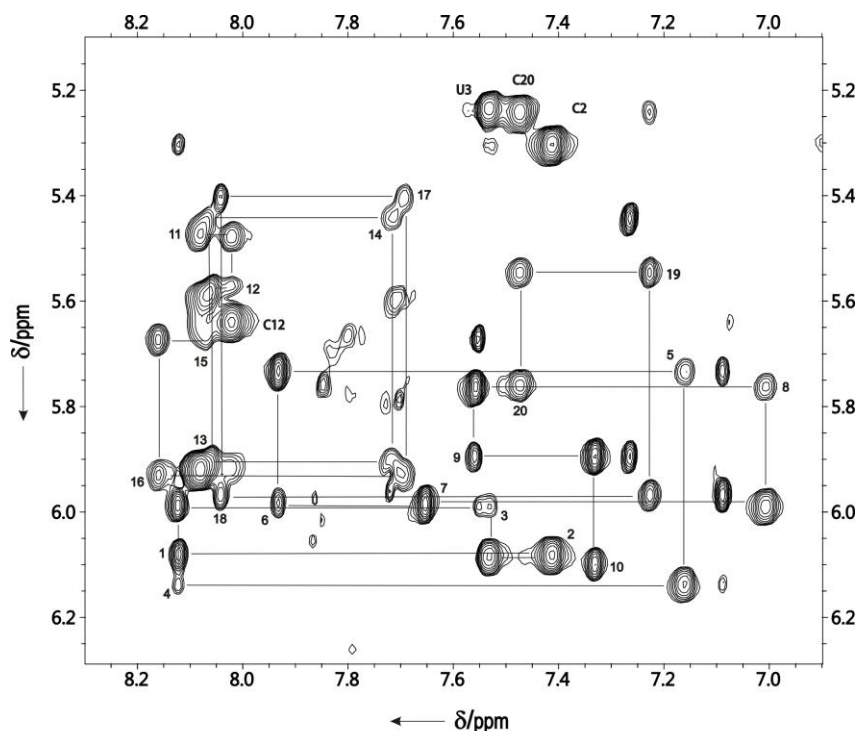


Figure 4: Region of the NOESY spectrum (250 ms mixing time) of the **AR** hybrid duplex in D_2O (0.8 mM oligonucleotide concentration, 25 mM sodium phosphate, 100 mM NaCl, $T=5^\circ C$, $pH=7$). Sequential assignment pathways for non-exchangeable protons are shown.

Exchangeable protons were assigned with the NOESY spectra recorded in H_2O (Figure 3). Most of the labile protons were assigned following standard methods, except some amino resonances of the terminal guanines that were not detected. The crosspeak patterns observed for the exchangeable protons indicate that all bases are forming Watson-Crick pairs throughout the duplex (Figure 3). Chemical shifts of exchangeable protons are almost identical in the **FR** and **DR** duplexes, and exhibit small changes in the **AR** duplex (Tables A6.1 (**FR**) and A6.2 (**AR**)). Assignment of ^{19}F resonances was carried out through their heteronuclear correlations with the adjacent $H2'$, $H3'$ and $H1'$ protons. Sequential and intra-residual ^{19}F – $^1H6/^1H8$ crosspeaks, along with three sequential ^{19}F –thymine methyl crosspeaks, were observed in the HOESY spectrum (Figure 5). Full assignments are given in Tables A6.1 (**FR**) and A6.2 (**AR**).

6.2.3. Experimental constraints and structure calculations.

Quantitative distance constraints were obtained from NOESY experiments by using a complete relaxation matrix analysis with the program MARDIGRAS (section 4.1.3.2). The total number of experimental distance constraints was around 200 (summarized in Table 2). Only structurally relevant distance constraints were included. Intra-residual constraints involving protons of the same sugar are not included, with the exception of H1'-H4' and H2"-H4'. Considering these constraints, the average number per base pair is around 20. In addition, a total of 21 $^{19}\text{F} - ^1\text{H}$ distance constraints were obtained from heteronuclear dipolar correlation experiments. Sequential HOE crosspeaks were particularly intense for the pyrimidine-purine steps (Figure 5).

Table 2: Summary of NMR constraints and structure calculations.

Experimental ^1H - ^1H distance constraints ^a	FR	AR
Total number	162	186
Intra-residue	61	80
Sequential	72	74
Inter-strand	29	32
RMSD ^b (Å)	FR	AR
all heavy atoms	0.7 ± 0.2	0.7 ± 0.2
backbone	0.9 ± 0.2	0.8 ± 0.2
all bases	0.5 ± 0.2	0.4 ± 0.1
Modified strand	0.6 ± 0.2	0.5 ± 0.1
RNA strand	0.7 ± 0.2	0.5 ± 0.1
sum of violations (Å) (average and range)	4.3 (4.0..4.8)	8.0 (7.7..8.2)
max. violation (average and range)	0.27 (0.2..0.3)	0.39 (0.3..0.4)
NOE energy kcal/mol (average and range)	19.3 (10..36)	31.5 (28..43)

^a 21 experimental ^{19}F - ^1H distance constraints were also used in the calculation of **FR**

^b Terminal residues are excluded from RMSD calculations

Some structural information can be readily determined from these distances. For example, intra-residual and sequential H1'-base and H2'-base NOEs in the RNA strand are consistent with a standard A-form duplex. However, many of the inter-proton distances in the 2'-F-ANA and ANA strands are not consistent with a pure A-

form conformation in the modified strand. This is particularly clear in the strong sequential H2"-H6/8 NOEs of the ANA strand in the **AR** hybrid, indicating significant B-like conformation in the ANA strand. In both 2'-F-ANA and ANA sugar moieties, H1'-H4' NOEs are strong, indicating a high population of an East-type sugar conformation, as expected^{1-3,8}.

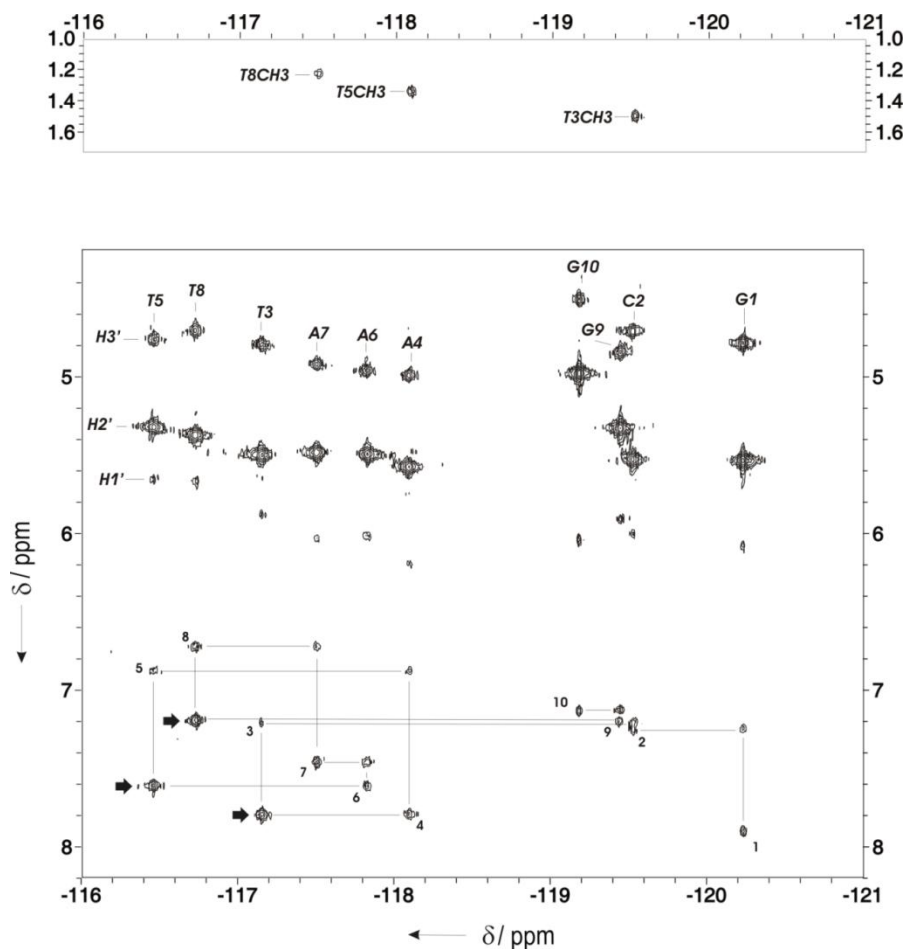


Figure 5: Heteronuclear $^{19}\text{F}\text{-}^1\text{H}$ correlation HOESY spectrum of **FR** in D_2O (0.8 mM oligonucleotide concentration, 25 mM sodium phosphate, 100 mM NaCl, $T=25^\circ\text{C}$, $\text{pH}=7$). Strong sequential $^{19}\text{F}(\text{T})\text{-H8}(\text{A})$ crosspeaks (at TA or TG steps) are indicated with arrows.

In addition to the NOE-derived information, a qualitative analysis of the J -coupling constants obtained from DQF-COSY spectra was carried out. $J_{1'2'}$ are undetectable in most of the riboses of the RNA strand of all the duplexes, indicating that riboses

are in a pure North-type conformation. In the 2'-F-ANA and ANA strands, however, $J_{1'2''}$ are medium or weak. This, together with the small $J_{3'2''}$ and the large $J_{4'3''}$ coupling constants (around 8-10 Hz), point towards a higher population of East conformation for these sugars.

According to this information, torsion angle constraints were set for the dihedral angles of the sugars. The riboses of the RNA strand were constrained to a North conformation. In the 2'-F-ANA and ANA strands, sugar dihedrals were constrained to avoid North conformations and allow the East and South regions. Backbone dihedral angles were not constrained. In case of 2'-F-ANA strand, heteronuclear J-couplings were detected between fluorine atoms and purine-H8 protons (Figure 6). These intra-residual couplings have been observed before in single 2'-F-ANA modified nucleotides and in 2'-F-ANA:2'-F-ANA duplex (chapter 5, section 5.2.3.)^{13,14}.

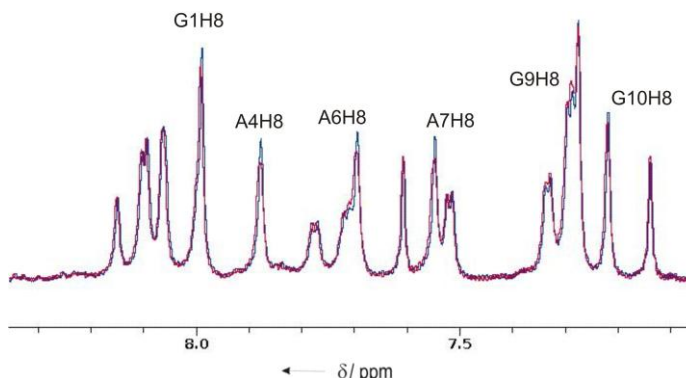


Figure 6: ^{19}F -coupled (red) and ^{19}F -decoupled (blue) ^1H NMR spectra, demonstrating different signal intensity in H8 2'-F-ANA purines. H8 of 2'-F-ANA purines in the duplex are labeled.

All the distance and torsion angle constraints were used to calculate the structure of both hybrid duplexes by restrained molecular dynamics as described in Annex I. The ten final structures of both hybrids resulting from restrained molecular dynamics calculation including the solvent explicitly are displayed in Figure 7. As can be seen in this Figure and also in Table 2, the two duplexes are well defined, with an RMSD of 0.7 Å (excluding the terminal residues). RMSD values are in the same range for the **FR** and **AR** duplexes. The final AMBER energies and NOE

terms are reasonably low in all the structures, with no distance constraint violation greater than 0.4 Å.

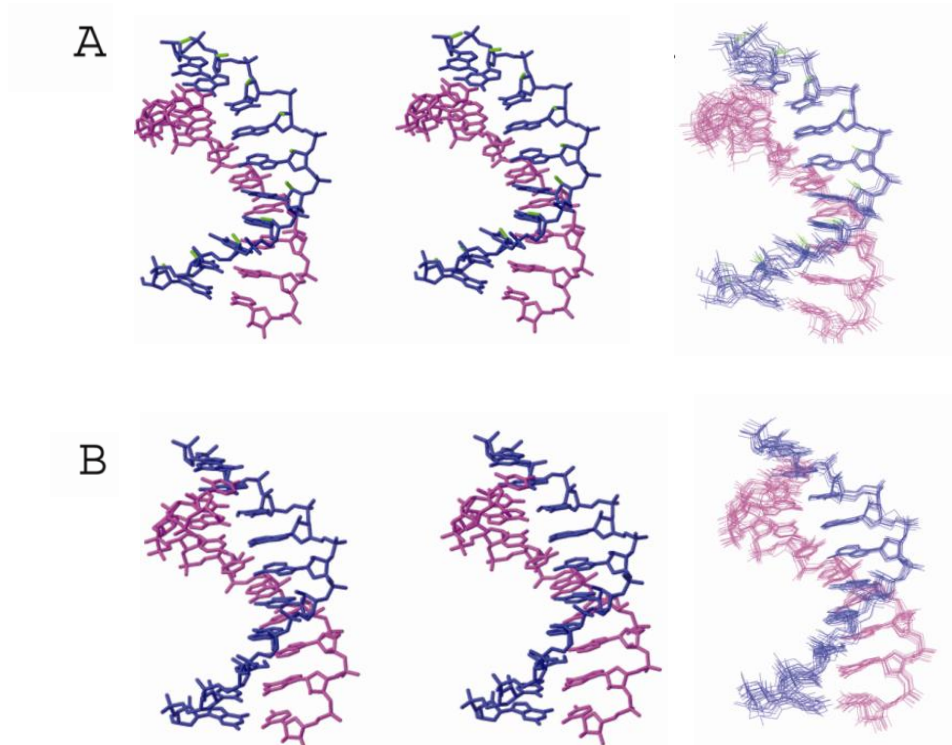


Figure 7: (A) Superposition of the 10 refined structures of **FR** (right) and stereoscopic views of the average structure (left). (B) Superposition of the 10 refined structures of **AR** (right) and stereoscopic views of the average structure (left).

6.2.4. Description of the structures

The overall shapes of **FR** and **AR** hybrids are intermediate between canonical A-form and B-form duplexes, but are closer to the A-form, as found for **DR** hybrids. The RMSD between the average structure of **FR** and **AR** hybrids and canonical A- and B-form duplexes are around 2.5 and 3.5 Å, respectively. RMSDs comparing the average structure of **FR**, **AR** and **DR** duplex are shown in Table 3. The RMSD between the **FR** and **AR** average structures is 1.5 Å for heavy atoms in the non-terminal residues. The deviation with the average structure of the control duplex is in the same range (1.8 Å).

Table 3: Root mean square deviation (Å) between **FR**, **AR** and **DR** hybrid duplex.

	FR	AR	DR
FR		1.5 1.0 1.2	1.8 1.6 1.2
AR	1.2 0.9 0.8		1.8 1.4 1.3
DR	1.4 0.9 1.2	1.3 1.2 1.1	

Upper right quadrant: all heavy atoms, lower left quadrant: base atoms. RMSD for both strands are in bold, for the modified strand in roman, and for the RNA strand in italics. Terminal residues are not included in the RMSD calculation.

The geometry of the RNA strand is very similar in the three hybrids, with most of the riboses in the C3'-*endo* conformation and glycosidic angles around -160° (the only exception is U14). Thus the RNA strand in hybrids is quite rigid and conformationally similar to the RNA homoduplex. In the **FR** and **AR** hybrids, pseudorotation phase angles of arabinoses and 2'-F-arabinoses are between 100° and 150°, in the southeast, and the glycosidic torsion angles range from -100° to -140°. These values tend to be higher than the corresponding ones for the deoxyriboses in the control duplex **DR**. This reflects an average between northern and southern conformations for **DR**, whereas for **FR** and **AR** the average occurs between eastern and southern puckers. Complete tables of geometrical parameters are shown in Tables A6.3 and A6.4.

The minor groove width in all three hybrids is intermediate between those of standard A-form and B-form helices (Figure 8). Of the three hybrids, however, **AR** features the narrowest minor groove width (Figure 8). This is consistent with a more B-like conformation for the **AR** hybrid.

Helical parameters are relatively dispersed and reflect that the hybrid structures are intermediate between A- and B-form. Whereas rise values are around 3.1 Å, typical of B-form helices, twist angles are around 32°, which is characteristic of A-form helices. Significant roll values are observed in all pyrimidine-purine steps in the **FR** duplex, probably related to the presence of 2'-F..H8 pseudohydrogen bonds (see below). A summary of important helical parameters is shown in Tables A6.5 and A6.6.

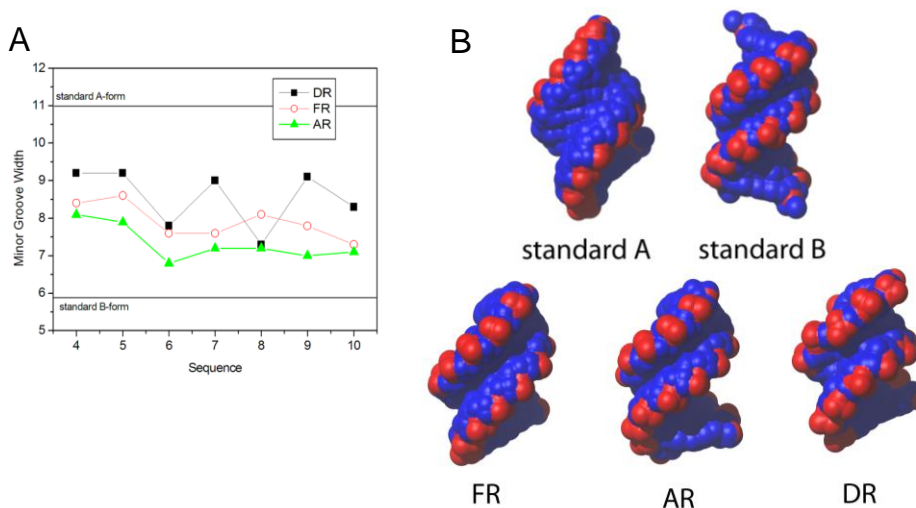


Figure 8: (A) Values of minor groove width along the sequence of the three hybrids, and (B) view of the hybrids from their minor groove side. Standard A-form and B-form duplexes are shown for comparison.

6.2.5. Fluorine chemical shifts

6.2.5.1. Hydration at fluorine atoms

NMR spectroscopy can give information about hydration at fluorine atoms by examining the chemical shift of fluorine nuclei in H₂O and D₂O buffers. For the **FR** duplex, the ¹⁹F chemical shift was relatively constant in both H₂O and D₂O (differences are lower than 0.01 ppm in all cases), implying that the fluorine atoms are not well hydrated. In contrast, nucleosides such as 2'-deoxy-2'-fluorocytidine and oligonucleotides containing solvent-exposed 2'-fluorine groups can show chemical shift changes up to 0.2 ppm¹⁵.

6.2.5.2. 2'F..H8 Pseudohydrogen bonds

Fluorine NMR signals of thymines involved in pyrimidine-purine pseudohydrogen bonds (T3, T5, T8) (see below) are downfield, and their chemical shift variation when temperature increases is different compared with the rest of fluorine signals (Figure 9). T3 and T5 fluorine signals don't move upon temperature, while T8

fluorine moves in the opposite direction compared with the rest of $2'\text{F}$ atoms. These effects have been also observed in “regular” hydrogen bonds. Normally, protons involved in the interaction behave different compared with the others¹⁶⁻¹⁸.

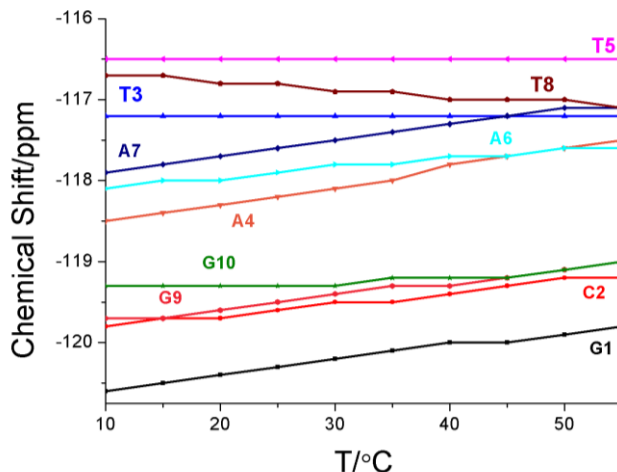


Figure 9: Fluorine signal chemical shift vs temperatures (T).

6.3. DISCUSSION

6.3.1. Structural differences

Despite the general similarity between **DR**, **AR** and **FR** hybrids, subtle structural differences can be observed between the modified strands. Thus, while sugars in the DNA strand of a **DR** hybrid are in a dynamic equilibrium between North and South (or A and B-form) conformations, sugars in both ANA and $2'\text{F}$ -ANA strands are more rigid and sampled only the South and East regions. $2'\text{F}$ -ANA sugars appear slightly more displaced towards the East region, while the sugars of the ANA strand are closer to the canonical B-form (South) region. Excluding terminal residues, the average pseudorotational phase angle is 125° for the $2'\text{F}$ -ANA strand and 133° for the ANA strand. Consistent with this increase in B-form character for the ANA strand, the minor groove width of **AR** is smaller than that of **FR** or **DR**. A third line of evidence for the greater B-form character of the ANA strand is found in the strong sequential $\text{H}2''$ - $\text{H}6/8$ NOEs observed in that strand. Our data are

consistent with early expectations that ANA nucleotides are unlikely to adopt the A-conformation¹⁹.

Since the RNA strand is maintained in a northern conformation, a strand with a rigid southern conformation is not easily tolerated, which might help justify the lower affinity of ANA to form hybrids with RNA. These subtle structural differences were not observed in the hairpin system^{1,2}, but a study on single inserts in A-form and B-form DNA duplexes concluded that ANA was restricted to the southeast pucker range, while 2'-F-ANA could adopt a much broader range of conformations including northeastern conformations¹¹. Our findings are also consistent with previous evidence from circular dichroism, which showed that the short wavelength (~210 nm) negative band associated with the A-form helical structure is slightly reduced in 2'-F-ANA•RNA hybrids and further reduced in ANA•RNA hybrids⁵.

6.3.2. Pseudohydrogen bonding and 2'-F..nucleobase interactions

The second clear difference that emerges between the 2'-F-ANA- and ANA-based hybrids is a bad steric interaction involving the 2'-OH group in **AR** which contrasts with favorable 2'-F..H8 (purine) pseudohydrogen bonds in the case of **FR**. Evidence for these interactions comes directly from different type of experiments. Intra-residual 2'-F..H8 interactions come from heteronuclear J-couplings detected between both atoms (Figure 6). This effect has been previously described in chapter 5 and in single 2'-F-ANA modified nucleotides^{13,14}. On the other hand, evidences for inter residue pseudo-hydrogen bonds in purine-pyrimidine steps come from HOESY NMR spectra (Figure 5), fluorine chemical shift displacement vs temperature (Figure 9), as well as the fact that the duplex geometry appears to be adjusted to optimize 2'-F..H8 interactions (Figure 10 and Table 4 for a listing of all intra and inter-residual 2'-F..H6/H8 distances).

Structural results reported here strongly suggest that optimal 2′F..H8 pseudohydrogen bonding is achieved at purine-pyrimidine steps, where the base-stacking geometry can adjust to optimize the interaction without incurring a steric penalty (Figure 10, also see Table 4). In purine-only 2′F-ANA sequences, this optimization may be harder to achieve, causing either reduced pseudohydrogen bonding or other structural problems, and ultimately leading to less stabilization as predicted²⁰. All these findings agree with the experimental observations^{5,21} that the highest increases in binding affinity upon replacement of a DNA strand by 2′F-ANA tend to occur for mixed base sequences.

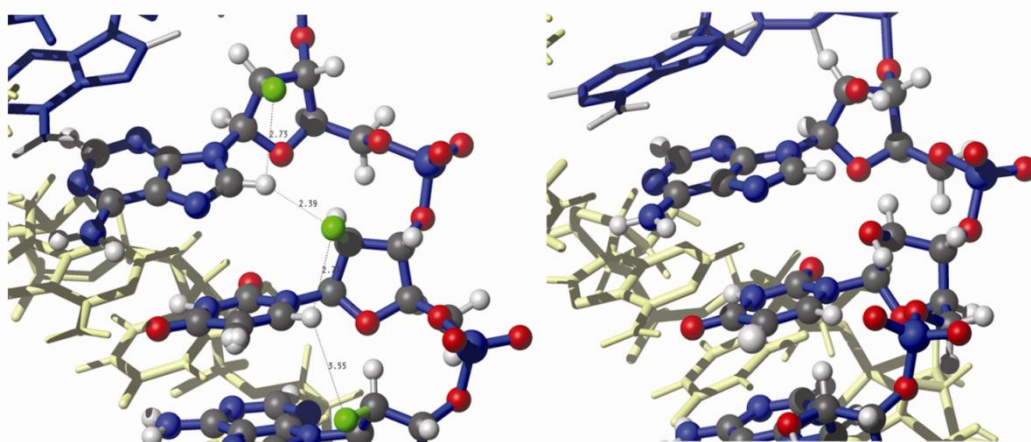


Figure 10: Details of X2′..H8 interactions in **FR** and **AR**. Left, structures of the **FR** duplex, showing distances between 2′F and aromatic H6/H8 protons. These distances correlate well with cross-peak intensity of the ¹⁹F-¹H HOESY spectra shown in Figure 5. Right, the same region of the **AR** duplex showing unfavourable steric interactions with 2′-OH.

The existence of pseudohydrogen bonds involving 2′F-ANA is consistent with previous experimental data. For example, single 2′F-ANA residues in a DNA helix showed short interresidue 2′F..C8 distances (<3 Å) in an A-form helical environment¹¹. Interestingly, this was also for a TA step (5′- 2′F-araT-dA -3′). If these interactions were stabilizing, it would help explain the significant distortion observed in the local helical environment around the residues¹¹. Short 2′F..C6 (pyrimidine) distances have also been observed in the crystal structure of a 2′F-ANA-modified B-form DNA sequence (2′F..C8 < 2.8 Å)²⁰.

Detailed energy calculation of the internal hybrid duplex interaction (i.e. intra-duplex energy) was calculated at Prof. Modesto Orozco's laboratory, and showed that the major reason for the increased stability of **FR** over **AR** is containing within the intra-duplex interaction energy, which is favorable for 2'F-ANA (better in the duplex than in the single strand) but unfavorable for ANA. A detailed analysis revealed that a sequential pseudohydrogen bond between 2'F..H8 contributes substantially to the stabilization of the **FR** duplex (around 2 kcal/mol), while bad steric contributions (differential van der Waals energy +0.5 kcal/mol) and the lack of favorable electrostatic contacts are responsible for the unfavorable intra-duplex binding energy at the **AR** derivative

Table 4: Intra- and interresidual 2'F–H6/8 distances for **FR** (Å). Short inter nucleotide distances (pyrimidine-purine steps) are labelled in red.

Intraresidual ($F_i \dots H6/H8_i$)			Interresidual ($F_i \dots H6/H8_{i+1}$)	
1	G	$2.3 \pm 0.1 \text{ \AA}$	G-C	$3.6 \pm 0.2 \text{ \AA}$
2	C	$2.5 \pm 0.2 \text{ \AA}$	C-T	$3.6 \pm 0.2 \text{ \AA}$
3	T	$2.5 \pm 0.1 \text{ \AA}$	T-A	$2.5 \pm 0.1 \text{ \AA}$
4	A	$2.3 \pm 0.1 \text{ \AA}$	A-T	$3.5 \pm 0.4 \text{ \AA}$
5	T	$2.8 \pm 0.1 \text{ \AA}$	T-A	$2.4 \pm 0.1 \text{ \AA}$
6	A	$2.8 \pm 0.4 \text{ \AA}$	A-A	$3.2 \pm 0.4 \text{ \AA}$
7	A	$2.4 \pm 0.1 \text{ \AA}$	A-T	$3.5 \pm 0.2 \text{ \AA}$
8	T	$2.6 \pm 0.1 \text{ \AA}$	T-G	$2.5 \pm 0.1 \text{ \AA}$
9	G	$2.7 \pm 0.3 \text{ \AA}$	G-G	$3.7 \pm 0.6 \text{ \AA}$
10	G	$2.5 \pm 0.3 \text{ \AA}$		

In contrast to the 2'F-ANA situation, 2'-OH..H8 (purine) interactions in **AR** are destabilizing by about 0.5 kcal/mol due to bad van der Waals contacts that are not compensated by favourable electrostatic contacts²². This quite surprising difference in the behaviour of 2'-OH and 2'F is justified by structural data which shows how the neighbouring phosphates impose conformational constraints to the 2'-OH groups, preventing them from adopting a more favourable orientation for interacting with H8 or other acidic groups in their vicinities. Support for our claims on the importance of OH-mediated steric repulsion in ANA hybrids can be found in previous studies that found that a single ANA insert in a B-form DNA duplex showed steric hindrance between the 2'-OH and the Me5 and C6 groups of the

neighboring thymine, and an associated computational study predicted that these steric effects would be more severe in the A-form¹⁹. Thus while the concepts of 2'-F..H8 mediated stabilization and 2'-OH..H8/C8 mediated destabilization are not totally unprecedented, we were surprised to find such strong evidence that they operate in the same sequence.

6.3.3. FC-H2''..O backbone interactions

Interestingly, in 2'-F-ANA:RNA duplex, short $H2''_i..O4'_{i+1}$ sequential contacts are observed. East sugar conformation of 2'-F-ANA nucleotides optimize these FC-H2''..O interactions, whereas southern conformations of ANA strand, in **AR** duplex, impede these contacts in the duplex. Therefore, these non conventional hydrogen bonds are also related with 2'-F-ANA:RNA stabilization compared with ANA:RNA and DNA:RNA. On chapter 7, we will prove the existence and the importance of FC-H..O interactions on the structure and stabilization of C2'-fluorine modified nucleic acids.

6.4. CONCLUSION

Several factors play a contributing role in the striking stability difference of DNA, ANA and 2'-F-ANA-containing hybrid duplexes. The most significant differences appear to be on the level of structure, steric contacts, and hydrogen bonding.

Both ANA and 2'-F-ANA-containing hybrids are structurally similar to DNA•RNA hybrid duplexes. The ANA strand is limited to more southern conformations, while the 2'-F-ANA strand is somewhat closer to the East, a conformation which is more compatible with the fixed northern conformation of the RNA strand. This is reflected, among other features, in a narrower minor groove width for the ANA•RNA hybrid duplexes.

Interactions between the 2'-substituent and the nucleobase are crucial in determining the different stabilities of **FR** and **AR**. For **FR**, several independent pieces of evidence, both computational and empirical, point to a favorable 2'-F..H8 (purine) and FC-H..O pseudohydrogen bonds. On the other hand, the 2'-OH..base interactions in the case of ANA destabilize the duplex due to unfavorable steric contacts.

Both ANA and 2'-F-ANA-based hybrids are more rigid than their DNA-containing counterparts. In principle this should benefit them both on an entropic level, but in practice any benefit conferred upon ANA is overwhelmed by the unfavorable structural characteristics discussed above. Furthermore, interactions between the 2'-substituent and the base may help preorganize the single strand into an appropriate conformation for duplex formation in the case of 2'-F-ANA, while disfavoring that conformation in the case of ANA.

The dramatic stability difference between these two very similar hybrids is therefore caused by the cumulative influence of multiple effects. Indeed, thermal stability can often be a surprisingly complex property²³⁻²⁵. While it is difficult to rationally design an oligonucleotide analogue to optimize so many factors at once, keeping these factors in mind can be informative as new analogues are developed. This study also beautifully illustrates the surprising effect of small changes by mutating the 2'-substituent of ANA from 2'-OH to 2'F, interactions with the nucleobases not only stopped being destabilizing, but started making a significant contribution to duplex stability.

BIBLIOGRAPHY

- (1) Trempe, J.-F.; Wilds, C. J.; Denisov, A. Y.; Pon, R. T.; Damha, M. J.; Gehring, K. *NMR Solution Structure of an Oligonucleotide Hairpin with a 2'-F-ANA/RNA Stem: Implications for RNase H Specificity toward DNA/RNA Hybrid Duplexes* **J. Am. Chem. Soc.** (2001), 123, 4896.
- (2) Denisov, A. Y.; Noronha, A. M.; Wilds, C. J.; Trempe, J.-F.; Pon, R. T.; Gehring, K.; Damha, M. J. *Solution structure of an arabinonucleic acid (ANA)/RNA duplex in a chimeric hairpin: comparison with 2'-fluoro-ANA/RNA and DNA/RNA hybrids* **Nucleic Acids Res.** (2001), 29, 4284.
- (3) Noy, A.; Luque, F. J.; Orozco, M. *Theoretical Analysis of Antisense Duplexes: Determinants of the RNase H Susceptibility* **J. Am. Chem. Soc.** (2008), 130, 3486.
- (4) Noy, A.; Pérez, A.; Márquez, M.; Luque, F. J.; Orozco, M. *Structure, Recognition Properties, and Flexibility of the DNA-RNA Hybrid* **J. Am. Chem. Soc.** (2005), 127, 4910.
- (5) Wilds, C. J.; Damha, M. J. *2'-Deoxy-2'-fluoro-6-d-arabinonucleosides and oligonucleotides (2'-F-ANA): synthesis and physicochemical studies* **Nucleic Acids Res.** (2000), 28, 3625.
- (6) C.J. Wilds, PhD Thesis, McGill University, 1999, Chapter 6.
- (7) Noronha, A. M.; Wilds, C. J.; Lok, C.-N.; Viazovkina, K.; Arion, D.; Parniak, M. A.; Damha, M. J. *Synthesis and Biophysical Properties of Arabinonucleic Acids (ANA): Circular Dichroic Spectra, Melting Temperatures, and Ribonuclease H Susceptibility of ANA-RNA Hybrid Duplexes* **Biochemistry** (2000), 39, 7050.
- (8) Watts, J. K.; Damha, M. J. *2'-F-Arabinonucleic acids (2'-F-ANA) — History, properties, and new frontiers* **Can. J. Chem.** (2008), 86, 641.
- (9) See the press release of December 22, 2008 at www.topigen.com.
- (10) Venkateswarlu, D.; Ferguson, D. M. *Effects of C2'-Substitution on Arabinonucleic Acid Structure and Conformation* **J. Am. Chem. Soc.** (1999), 121, 5609.
- (11) Li, F.; Sarkhel, S.; Wilds, C. J.; Wawrzak, Z.; Prakash, T. P.; Manoharan, M.; Egli, M. *2'-Fluoroarabino- and Arabinonucleic Acid Show Different Conformations, Resulting in Deviating RNA Affinities and Processing of Their Heteroduplexes with RNA by RNase H* **Biochemistry** (2006), 45, 4141.

- (12) Gonzalez, C.; Stec, W.; Reynolds, M. A.; James, T. L. *Structure and Dynamics of a DNA.RNA Hybrid Duplex with a Chiral Phosphorothioate Moiety: NMR and Molecular Dynamics with Conventional and Time-Averaged Restraints* **Biochemistry** (1995), 34, 4969.
- (13) Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; Gonzalez, C. *The solution structure of double helical arabino nucleic acids (ANA and 2'-F-ANA): effect of arabinoses in duplex-hairpin interconversion* **Nucleic Acids Res** (2012), 40, 9329.
- (14) Anzahaee, M. Y.; Watts, J. K.; Alla, N. R.; Nicholson, A. W.; Damha, M. J. *Energetically Important C-H...F-C Pseudohydrogen Bonding in Water: Evidence and Application to Rational Design of Oligonucleotides with High Binding Affinity* **J. Am. Chem. Soc.** (2010), 133, 728.
- (15) Fenna, C. P.; Wilkinson, V. J.; Arnold, J. R.; Cosstick, R.; Fisher, J. *The effect of 2'-fluorine substitutions on DNA i-motif conformation and stability* **ChemComm** (2008), 3567.
- (16) Cierpicki, T.; Otlewski, J. *Amide proton temperature coefficients as hydrogen bond indicators in proteins* **J. Biomol. NMR** (2001), 21, 249.
- (17) Cierpicki, T.; Zhukov, I.; Byrd, R. A.; Otlewski, J. *Hydrogen Bonds in Human Ubiquitin Reflected in Temperature Coefficients of Amide Protons* **J. Magn. Reson.** (2002), 157, 178.
- (18) Baxter, N. J.; Williamson, M. P. *Temperature dependence of ¹H chemical shifts in proteins* **J. Biomol. NMR** (1997), 9, 359.
- (19) Gao, Y. G.; Van der Marel, G. A.; Van Boom, J. H.; Wang, A. H. J. *Molecular structure of a DNA decamer containing an anticancer nucleoside arabinosylcytosine: conformational perturbation by arabinosylcytosine in B-DNA* **Biochemistry** (1991), 30, 9922.
- (20) Berger, I.; Tereshko, V.; Ikeda, H.; Marquez, V. E.; Egli, M. *Crystal structures of B-DNA with incorporated 2'-deoxy-2'-fluoro-arabino-furanosyl thymines: Implications of conformational preorganization for duplex stability* **Nucleic Acids Res.** (1998), 26, 2473.
- (21) Tennilä, T.; Azhayeva, E.; Vepsäläinen, J.; Laatikainen, R.; Azhayev, A.; Mikhailopulo, I. A. *Oligonucleotides containing 9-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)-adenine and-guanine: synthesis, hybridization and antisense properties* **Nucleos. Nucleot. Nucl.** (2000), 19, 1861.
- (22) Watts, J. K.; Martín-Pintado, N.; Gómez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; González, C.; Damha, M. J. *Differential stability of 2'-F-ANA•RNA and ANA•RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility* **Nucleic Acids Res.** (2010), 38, 2498.
- (23) Freier, S. M.; Altmann, K.-H. *The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA: RNA duplexes* **Nucleic Acids Res.** (1997), 25, 4429.
- (24) Egli, M.; Minasov, G.; Tereshko, V.; Pallan, P. S.; Teplova, M.; Inamati, G. B.; Lesnik, E. A.; Owens, S. R.; Ross, B. S.; Prakash, T. P. *Probing the influence of stereoelectronic effects on the biophysical properties of oligonucleotides: comprehensive analysis of the RNA affinity, nuclease resistance, and crystal structure of ten 2'-O-ribonucleic acid modifications* **Biochemistry** (2005), 44, 9045.
- (25) Kolarovic, A.; Schweizer, E.; Greene, E.; Gironda, M.; Pallan, P. S.; Egli, M.; Rozners, E. *Interplay of structure, hydration and thermal stability in formacetal modified oligonucleotides: RNA may tolerate nonionic modifications better than DNA* **J. Am. Chem. Soc.** (2009), 131, 14932.

Chapter 7

Stabilizing FC-H..O electrostatic interactions in 2'F-substituted Nucleic Acids

RESUMEN

En este capítulo estudiamos dos dúplex quiméricos con secuencias continuas y alternadas de 2'-F-ANA y 2'-F-RNA en su secuencia. En las regiones con tramos continuos de cada una de las modificaciones, 2'-F-ANA adopta una conformación sur/este y 2'-F-RNA una conformación norte. Sin embargo, en secuencias alternas, 2'-F-RNA adopta una conformación inusual tipo este. La polarización de carga inducida por el flúor en posición 2' provoca interacciones favorables con los oxígenos circundantes. En las regiones con el mismo tipo de modificación, la geometría del nucleótido hace que este efecto de polarización fortalezca enlaces de hidrógeno no convencionales del tipo FC-H2'/H2''..O4'. Estas interacciones secuenciales son, en parte, responsables de la mayor estabilidad de los dúplex de 2'-F-RNA con respecto a los de RNA. La ruptura de esta interacción en algunos pasos 2'-F-RNA/2'-F-ANA es la causa de la menor estabilidad de las secuencias alternas frente a las que contienen tramos continuos de cada una de las modificaciones.

ABSTRACT

In this chapter, we study two chimeric duplexes containing either continuous tracts or alternated sequences of 2′F-ANA and 2′F-RNA nucleotides. In regions containing continuous tracts, 2′F-ANA sugars adopt South/East-type conformations, and 2′F-RNA, North conformations. However, 2′F-RNA adopts an unusual East-type conformation in alternated sequences. The positive charge polarization of the proton at 2′-position induced by the fluorine provokes favorable interactions with the surrounding oxygen atoms. In tracts of the same type of modification, the nucleotide geometry provokes that this charge polarization effect strengthens non conventional hydrogen bond FC-H2′/H2′′..O4′. Such sequential interactions are, in part, responsible of the enhanced stability of 2′F-RNA over pure RNA duplexes. The disruption of these sequential interactions in some 2′F-RNA/2′F-ANA steps is the reason of the lower stability of alternate vs continuous 2′F-substituted sequences.

7.1. INTRODUCTION

Stabilization of nucleic acids secondary structures results from a subtle balance of multiple interactions. Some of them, like base pairing, base stacking or cation binding, have been extensively studied for decades¹⁻³. However, other effects still remain poorly understood. Among them, non conventional C-H...O hydrogen bonds are especially interesting. The importance of these interactions in proteins was recognized more than 40 years ago^{4,5} and, during the last few years, multiple studies have shown their relevance in multiple areas⁶⁻¹¹. In nucleic acids, most studies have focused on aromatic C-H...O interactions and, in particular, on the effect of nonconventional bonds in base pair strength and in base-sugar interactions¹⁰⁻¹². Backbone C-H...O hydrogen bonds have received less attention, although it has been proposed to play a role in stabilization of four stranded i-motif structures¹³ and in tight packing of RNA and DNA elements¹⁴.

Although nonconventional hydrogen bonds are considered weak interactions in natural nucleic acids, their strength can be increased through chemical modifications. In this regard, fluorine substitutions are particularly relevant since fluorine is the most electronegative element and can provoke significant charge polarization effects in geminal and vicinal protons, which increases the strength of nonconventional C-H...O hydrogen bonds. Such an effect has been well-characterized in small molecular systems^{15,16}. In a very recent report, Kakshoor *et al.* have observed fluorine-enhanced aromatic CH...N interactions that significantly affect the strength of 2,4-difluorotoluene–adenine base pairs¹⁷. These results prompted us to explore the occurrence of other fluorine enhanced nonconventional C-H...O interactions in fluorine-modified oligonucleotides.

2'-Deoxy-2'-fluoroarabinose (2'-F-ANA) and 2'-deoxy-2'-fluororibose (2'-F-RNA) are two widely used oligonucleotide chemical modifications containing fluorine substitution. Both modifications feature enhanced binding affinity for complementary RNA, but contrast with respect to their conformational preferences. 2'-F-ANA closely mimics DNA^{18,19}, and favors a South/East sugar conformation²⁰. On the other hand, 2'-F-RNA is an RNA mimic which tends to adopt a C3'-*endo* conformation. Indeed, fluorine substitutions at the C2' sugar position provoke

intriguing effects in nucleic acid stability that have attracted considerable attention and motivated numerous structural studies recently¹⁹⁻²¹.

Adding to our interest in the structural consequences of 2'-F-ANA and 2'-F-RNA modification, numerous studies have revealed that these nuclease resistant nucleic acid analogues are able to trigger, respectively, RNase H - and RISC - mediated RNA degradation. For these reasons, 2'-F-ANA and 2'-F-RNA are very promising modifications in antisense and siRNA therapies. Pure 2'-F-RNA and chimeric 2'-F-RNA:RNA duplexes adopt A-form structures with increased thermal stability with respect to native RNA duplexes. Recent studies indicate that this higher stability is mainly due to enhanced stacking interactions between 2'-F-RNA nucleobases and to an increased hydrogen bond strength between 2'-F-RNA base pairs²².

Interestingly, although 2'-F-ANA is a DNA-mimic, it is also compatible and often stabilizing within a dsRNA context. Very recently it has been shown that combining 2'-F-ANA and 2'-F-RNA in siRNA duplexes can provoke gene silencing effects of higher potency than pure RNA:RNA duplexes²³. An interesting observation is that, in spite of the high binding affinities of 2'-F-RNA and 2'-F-ANA substitutions, sequence combinations of both nucleotides decrease the stability of the duplex if combined in an alternating 1-1 pattern^{18,23}. This observation is difficult to explain solely on the basis of nucleobase polarization provoked by 2'-F substitutions, and prompted us to determine the structure of chimeric duplexes with different combinations of 2'-F-ANA and 2'-F-RNA in their sequence.

Here, we report an NMR and computational study of two chimeric self-complementary duplexes, containing continuous tracts (duplex **H**) and alternating regions (duplex **S**) of 2'-F-ANA and 2'-F-RNA substitutions (Figure 1). As demonstrated below, these duplexes are excellent model systems to explore the effect of different orientations of 2'-F along the sugar-phosphate backbone. The analysis of these and other crystallographic structures and quantum mechanical calculations clearly support the existence of sequential FC-H2'_i...O4'_{i+1} hydrogen bonds that are particularly strong in neighboring 2'-F-RNA residues. Although these interactions are present in pure RNA duplexes, the higher electron withdrawing

character of fluorine polarizes the C-H2' bond, significantly increasing the strength of this nonconventional C-H...O interaction. The lack of these interactions in 2'F-RNA_{*i*}/2'F-ANA_{*i+1*} steps helps explain, at least in part, the destabilization induced by alternating 2'F-ribose and 2'F-arabinose units along the duplex.

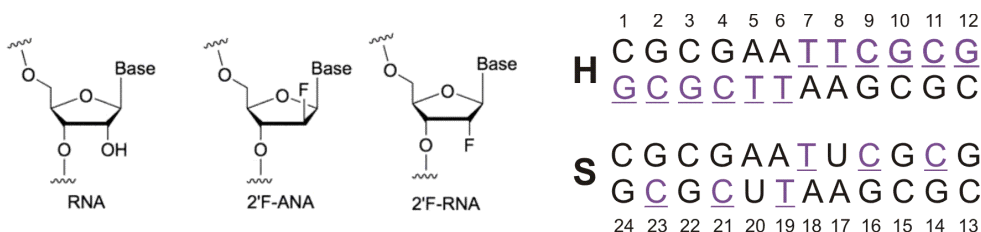


Figure 1: Structures of 2'F-ANA and 2'F-RNA in comparison with RNA (left). Numbering scheme of the 12-mer chimeric duplexes (**H** and **S**) used for this study (right). 2'F-ANAs are underlined in purple and 2'F-RNAs are in black.

7.2. RESULTS AND DISCUSSION

7.2.1. Duplex formation and thermal stability

Self-complementary duplex formation and melting was monitored by UV absorbance at 260 nm and NMR. Transitions for both samples, **H** and **S**, were reversible indicating a true equilibrium process. UV melting experiments show that **H** is 6° less stable than **FF** and more than 15°C than **rFrF** duplex, which exhibits the higher thermal stability. In contrast, **H** is around 20° more stable than **S** and the unmodified DNA sequence, and 5° more stable than the pure RNA duplex. T_m values are shown in Table 1.

Table 1: T_m values and thermodynamic parameters for the self-complementary chimeric duplexes (buffer conditions: 140 mM KCl, 1 mM $MgCl_2$ and 5 mM Na_2HPO_4 , pH 7.2).

Name	Sequence (5'-3') ^a	T_m (°C)
DD	5'-CGCGAATTCGCG-3'	58.4
RR	5'-CGCGAATTCGCG-3'	65.1
FF	5'-CGCGAATTCGCG-3'	76.1
rFrF	5'-CGCGAATTCGCG-3'	> 85
H	5'-CGCGAATTCGCG-3'	70.4
S	5'-CGCGAATUCGCG-3'	48.8

^a Legend: 2'-F-ANA, RNA, DNA and 2'-F-RNA.

¹H NMR spectra of **H** and **S** clearly indicate self-complementary duplex formation in both cases, with formation of Watson-Crick base pairs as shown by the cross-peak pattern in the NOESY spectra (Figure 2).

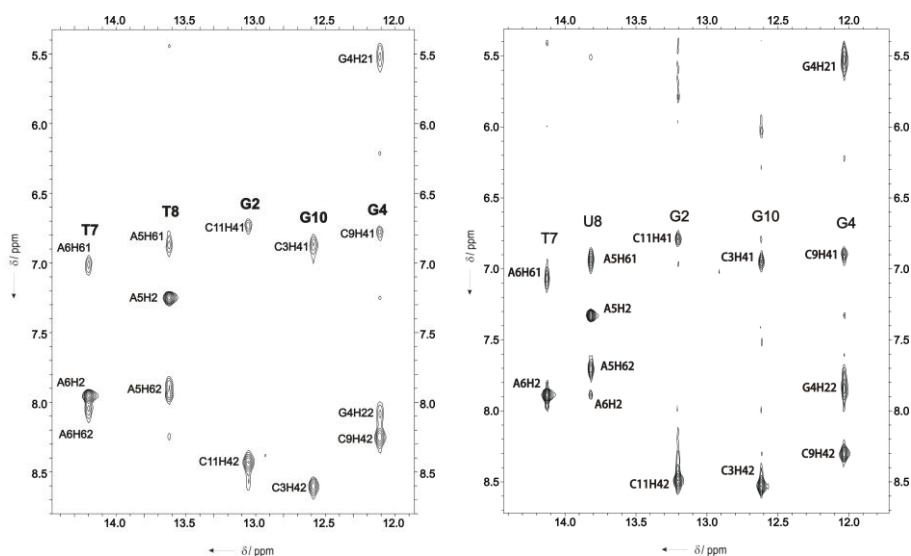


Figure 2: Imino region of the NOESY spectra in H_2O ($\tau_m=50$ ms) of **H** (left), and **S** (right). Buffer conditions were 100 mM NaCl, 25 mM sodium phosphate, $T=10$ °C, pH=7.

^1H and ^{19}F NMR melting experiments shown in Figures 3 and 4 confirm the higher stability of **H** vs **S** observed in UV melting experiments.

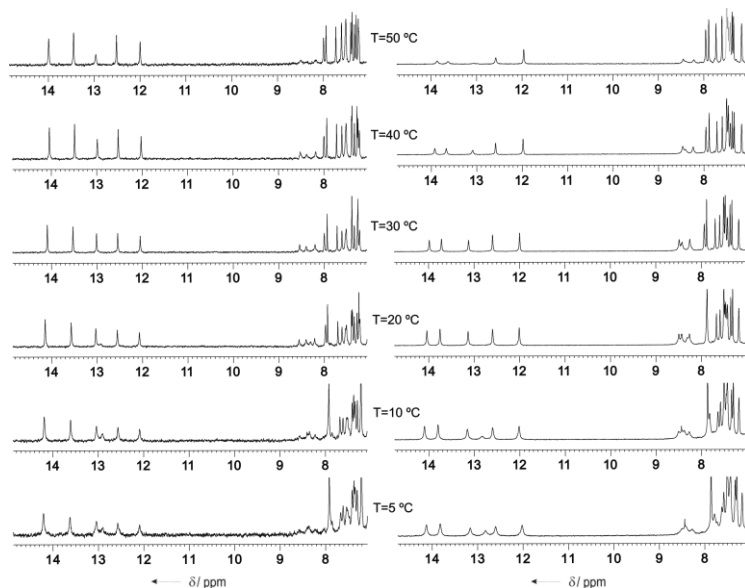


Figure 3: Left) ^1H spectra of **H** at different temperatures. Right) ^1H spectra of **S** at different temperatures. Oligonucleotide concentration 0.7 mM (25 mM sodium phosphate, 100mM NaCl, pH=7, $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1).

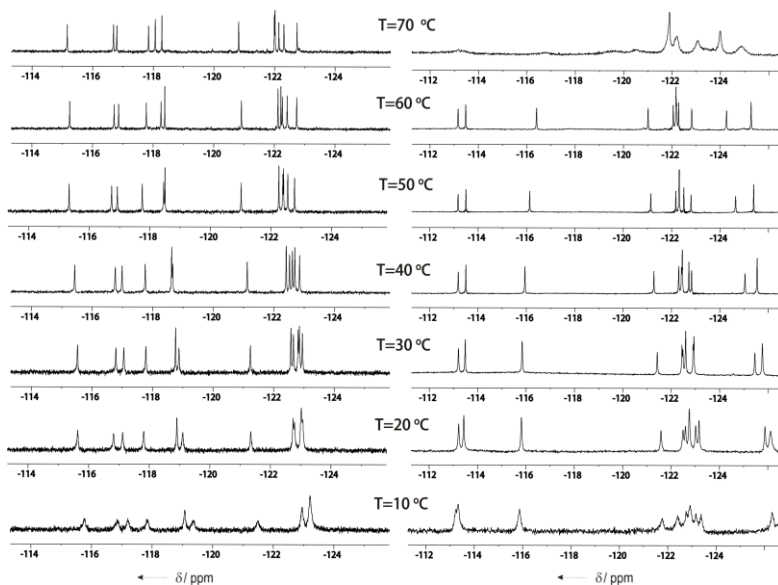


Figure 4: Complete ^{19}F spectra of **H** (left) and **S** (right) at different temperatures.

7.2.2. Experimental NMR data and structure calculations

Sequential assignments of ^1H resonances were performed following standard methods for right-handed double-stranded nucleic acids. Assignment of ^{19}F resonances was carried out through their correlations with the adjacent $\text{H}2'/\text{H}2''$, $\text{H}3'$, $\text{H}4'$ and $\text{H}1'$ protons in HOESY and HETCOR experiments (Figure 5). Sequential $^{19}\text{F}_i\text{-H}6/\text{H}8_{i+1}$ HOE cross-peaks, but not intra-residual $^{19}\text{F}_i\text{-H}6/\text{H}8_i$, were observed in $2'\text{F}$ -RNA continuous tracts and in $2'\text{F}$ -RNA/ $2'\text{F}$ -ANA steps of **S** and **H**. $2'\text{F}$ -ANA sequential and intra-residual ^{19}F -H6/H8 HOEs were observed in both structures. In A6/T7 in **H** and **S** these cross-peaks were not observed.

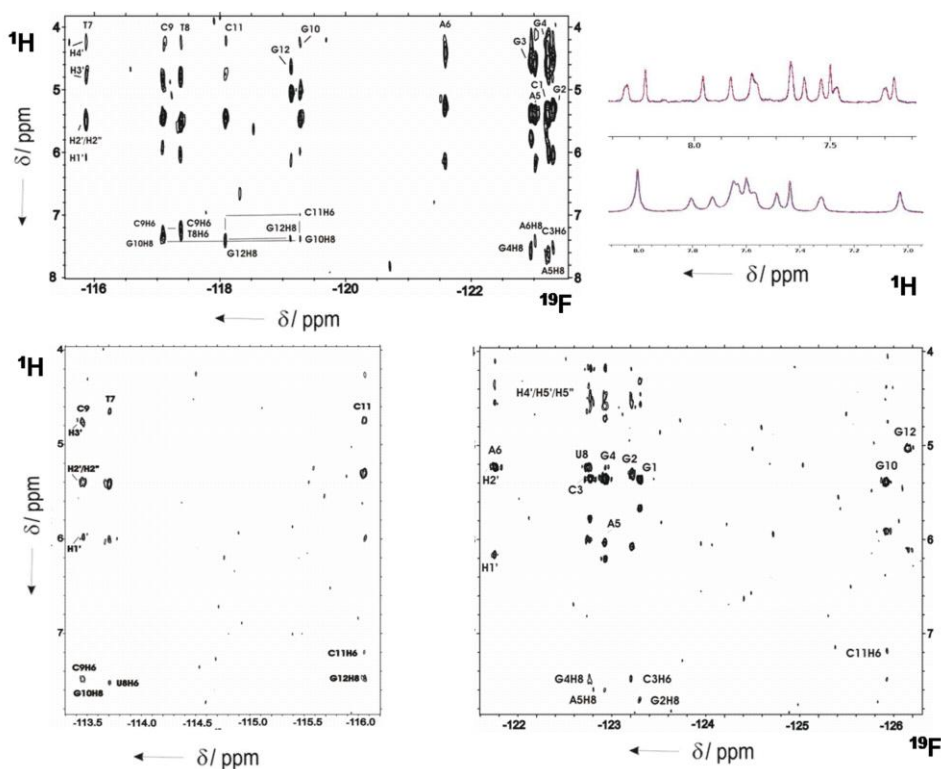


Figure 5: Heteronuclear ^{19}F - ^1H correlation HOESY spectrum of **H** (top left) and **S** (bottom) in D_2O (sample conditions: 100mM NaCl, 25 mM sodium phosphate, $T=25^\circ\text{C}$, $\text{pH}=7$). ^{19}F -coupled (red) and ^{19}F -decoupled (blue) ^1H NMR spectra of **H** and **S** (top right). Same intensity in all H6/H8 signals.

Complete assignment lists are given in Supplementary Table A7.1 and A7.2. A total number of 254 and 268 experimental proton-proton distance constraints were obtained from the NOESY experiments of **H** and **S**, respectively, plus about 60 fluorine-proton distances obtained from HOESY experiments (Table 2). In the continuous 2′F-RNA tracts of **H** and **S**, $J_{1'2'}$ are very weak and almost undetectable, indicating pure North-type conformations (Figure 6 left). However, 2′F-RNAs located in alternate 2′F-RNA/2′F-ANA steps (**S**) exhibit very different coupling constants (Figure 6, middle), which do not correspond with pure North type conformation, but with a significant population of East-South conformation. In the case of 2′F-ANA nucleotides, intermedium $J_{1'2'}$ and small $J_{3'2'}$ indicate a South/East conformation in **H** and **S** (Figure 6 right). According to this information, sugar torsion angle constraints were set to North for 2′F-RNAs in continuous tracts, while 2′F-RNA residues in alternate 2′F-RNA/2′F-ANA steps were not constrained. 2′F-ANAs sugar dihedral angles were loosely constrained to avoid North conformation and allow East and South regions.

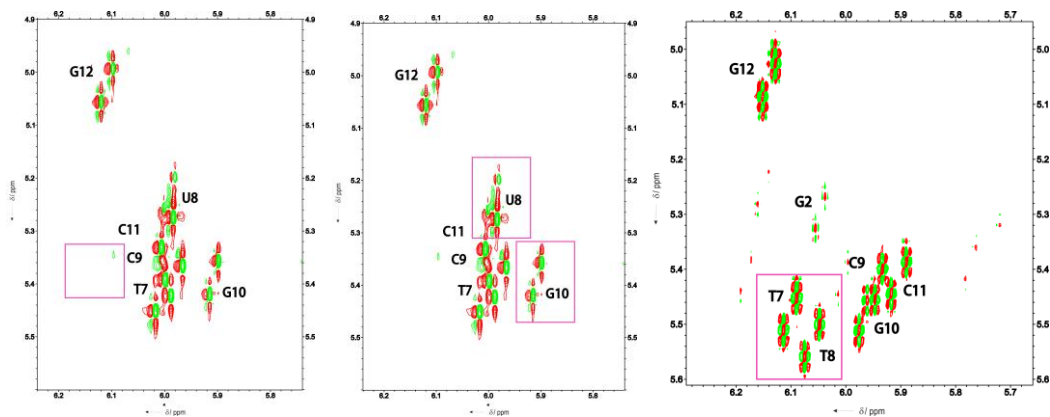


Figure 6: Region of the DQF-COSY spectrum showing $H2'/H2''-H1'$ cross-peaks of **H** and **S** in D_2O (100 mM NaCl, 25 mM sodium phosphate, $T=25^\circ C$, $pH=7$). 2′F-RNA weak (left) and medium (middle) signals are indicated by purple squares in **S** sample. 2′F-ANA (right) medium intensity signals are indicated by a purple square in **H** sample.

Distance and torsion angle constraints were used to calculate the structure by restrained molecular dynamics as described in Annex I. The ten resulting

structures are displayed in Figure 7. In both cases, the calculations converge to a well-defined ensemble, with an RMSD of 0.9 Å for heavy atoms and 0.7 Å for bases (excluding the terminal residues). Structure calculation statistics are given in Table 2.

Table 2: NMR restraints and structural statistics. The final AMBER energies and NOE terms are reasonably low in all the structures, which do not exhibit significant constraint violations.

Experimental ^1H - ^1H and ^{19}F - ^1H distance constraints	H	S
Total number	254	268
Intra-residue	122	118
Sequential	76	83
Inter-strand	4	4
^{19}F - ^1H	52	63
RMSD⁺ (Å)		
all heavy atoms	0.9 ± 0.2	0.9 ± 0.2
backbone	1.0 ± 0.3	1.1 ± 0.3
all bases	0.7 ± 0.2	0.7 ± 0.2
sum of violations (Å) (average and range)	5.1 (4.7-5.4)	4.7 (4.2-5.5)
max. violation (average and range)	0.7 (0.7-0.8)	0.6 (0.6-0.7)
NOE energy kcal/mol (average and range)	53 (51-56)	39 (37-43)

⁺ All except residues terminal residue

7.2.3. Structural properties of H and S duplexes

The structures of **H** and **S** are intermediate between canonical A and B-forms and retain structural features of both families of double helical structures. RMSDs between both structures and pure A and B-form duplexes are around 3 Å and 5 Å, respectively. The RMSD between **H** and **S** is 1.2 Å for all heavy atoms, excluding terminal residues.

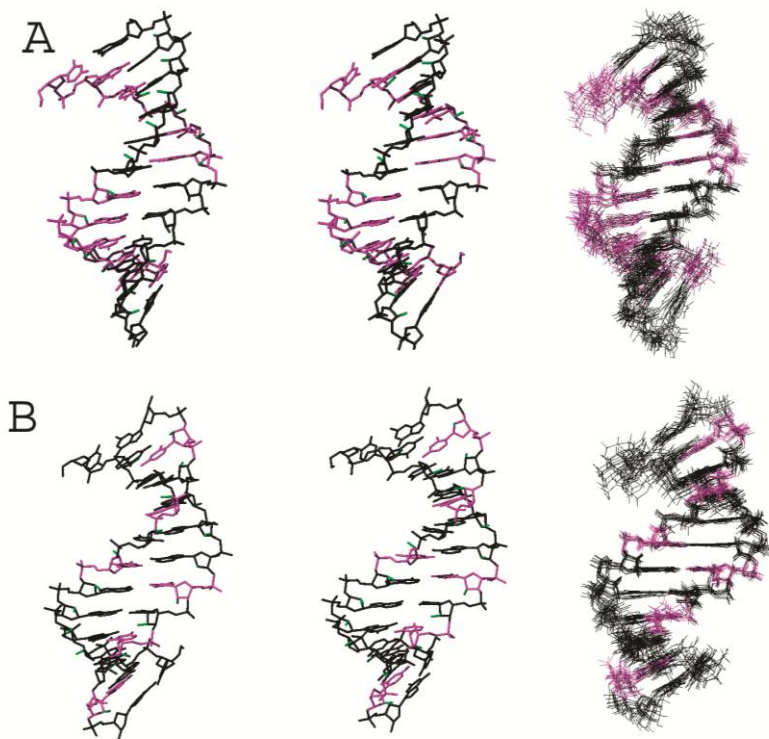


Figure 7: Stereoscopic views of the average structures and ensemble of the two chimeric duplexes **H** (A) and **S** (B). 2'-F-RNA nucleotides are shown in black, 2'-F-ANA nucleotides in magenta and fluorine atoms in green.

The geometry of 2'-F-RNA nucleotides in the continuous tracts of **H** and **S** is very similar to that expected for a RNA homoduplex, with riboses adopting a C3'-*endo* conformation (phase angles around 40-60°), and glycosidic torsion angles between -150° and -170°, except for terminal residues in **S** (Figure 8). Interestingly, 2'-F-ribose conformations of alternate 2'-F-RNA/2'-F-ANA steps in **S** are within the East domain, with pseudorotation phase angles between 80° and 95°, and glycosidic torsion angles ranging from -140° to -150° (Figure 8). This unexpected ribose conformation in the steps region is fully consistent with the experimental J-coupling data described above. The 2'-F-arabinoses adopt South/East conformation, with pseudorotation phase angles between 75° and 140° in **H**, and 100° to 160° in **S**. Geometrical parameters are shown in Tables A7.3-A7.4. Some helical parameters

present values typical of B-form, i.e. rise around 3.5 Å, and others are characteristic of A-form, i.e., twist around 27°. The minor groove width in both duplexes is intermediate between that of standard A and B-forms.

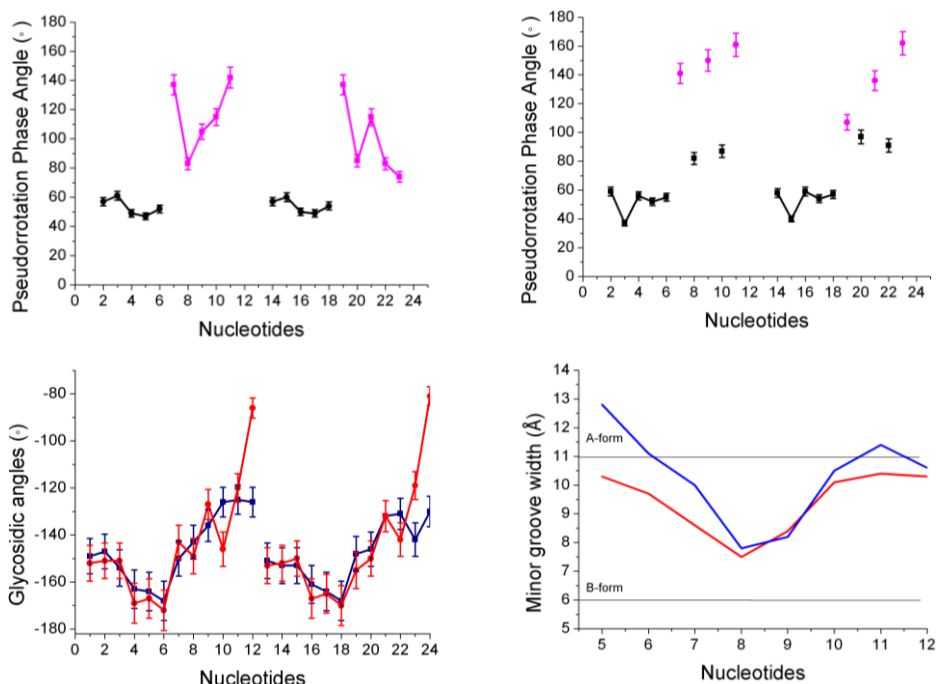


Figure 8: Top) Pseudorotation phase angle vs sequence in **H** (left) and **S** (right). 2'F-RNA nucleotides are in black and 2'F-ANA nucleotides in magenta. Bottom) Glycosidic angle vs sequence (left) and minor groove width vs sequence (right). **H** values are in blue, and **S** values in red.

7.2.4. H and S duplexes are devoid of 2'F..H8 pseudohydrogen bonds

Since 2'F-ANA and 2'F-RNA nucleotides are considered, respectively, as DNA and RNA analogs, it is interesting to compare the structures of the 2'F-ANA:2'F-RNA segments of **H** with that of 2'F-ANA:RNA hybrids. In the structure of a 2'F-ANA:RNA hybrid decamer (chapter 6 and PDB ID: 2KP4)²⁰, 2'F-ANA nucleotides adopt an East sugar conformation and glycosidic angles between 110° and 120°. Such geometry provokes a close 2'F..H8 contact, with a 2'F...H8-C angle around

150° that provides a very favorable geometry for pseudohydrogen bond formation²⁰. In contrast, glycosidic angles of contiguous tracts of 2'F-ANA residues in **H** are between 130° and 140° (Figure 8). These higher anti values in **H** are typical of a pure A-duplex and prevent the appropriate orientation for 2'F...H8 pseudo-hydrogen bond formation. The experimental evidence for this interaction in 2'F-ANA:RNA hybrids arise from strong sequential 2'F-H8 HOE cross-peaks, absence of fluorine chemical shift displacement upon temperature and heteronuclear ¹⁹F-¹H J-couplings for inter- and intra-interactions, respectively. None of these experimental data are observed in the 2'F-ANA:2'F-RNA chimeric duplexes studied here (Figure 5).

7.2.5. The effect of FC-H..O backbone interactions

Very close contacts between H2'/H2'' protons and their 3'-neighbour O4'-sugar and O5'-backbone atoms are observed in the continuous tracts of 2'F-ANA or 2'F-RNA (Figure 9 and Tables A7.5-A7.6). The H2'/H2''...O distances are particularly short in 2'F-ANA tracts (around 2.8 Å excluding terminal residue), where the C-H2''_i..O4'_{i+1} angles are close to co-linearity (average of 160°, see Table A7.5).

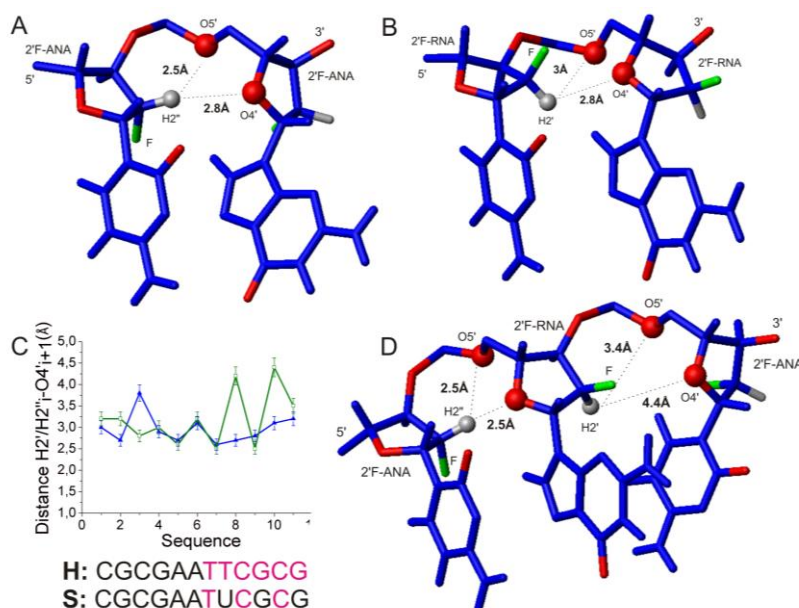


Figure 9: A) View of short distances between $H2''_{i-1} \cdots O4'/O5'_{i+1}$ oxygens and $H2'_{i-1} \cdots O4'/O5'_{i+1}$ in 2'-F-ANA and B) 2'-F-RNA continuous tracts of **H** and **S** respectively. C) Values of $H2'/H2'' \cdots O4'$ distances all along the sequence (F-RNA in black and 2'-F-ANA in magenta) in **H** (blue) and **S** (green). D) Structure and view of distances in steps of **S**. Oxygens are shown in red, protons in grey and fluorines in green.

Due to the high fluorine electronegativity, the geminal $H2'/H2''$ in 2'-F-ANA and 2'-F-RNA nucleotides are positively polarized, provoking a favourable electrostatic interaction with the surrounding electron dense oxygen atoms in the oligonucleotide backbone (Figure 9). In particular, the $C-H_{i-1} \cdots O4'_{i+1}$ geometry (Table 3) resembles that of nonconventional C-H...O hydrogen bonds observed in different contexts (for a recent review see Horowitz *et al*, 2012)⁸. In our case, the positive charge polarization of $H2'/H2''$ protons induced by the geminal 2'-F is supported by their NMR chemical shifts, which in 2'-F-ANA and 2'-F-RNA exhibit significantly higher values (5-6 ppm) than those typically observed in RNA (4 ppm).

The dramatic impact of these sequential C-H...O interactions on duplex stability becomes apparent when the structures of the **H** and **S** duplexes are compared. The favourable $H2'_{i-1} \cdots O4'_{i+1}$ contacts (2.9 Å) observed in the continuous tracts are

disrupted in some of the 2'F-RNA_{*i*}/2'F-ANA_{*i+1*} junctions (8→9 and 10→11) of duplex **S** (H2'_{*i*}...O4'_{*i+1*} distance around 4.4 Å) (see Figure 9 and Table 3). These distances are closely connected with the sugar conformations in 2'F-ANA/2'F-RNA and in 2'F-RNA/2'F-ANA junctions. Alternation of South 2'-fluoroarabinoses and East 2'-fluororiboses favours C2'-H...O4' contacts in 2'F-ANA_{*i*}/2'F-RNA_{*i+1*} steps, and prevent their formation in the adjacent 2'F-RNA_{*i*}/2'F-ANA_{*i+1*} step (see Figure 9D). Comparing **H** with **S**, the **S** duplex lacks two FC-H2'...O4' contacts present in **H**, and the ΔT_m between the two is 21.6 °C. The missing FC-H2'...O4' electrostatic interactions could help explain this large difference in stability.

The East 2'F-RNA conformation, provoked by neighboring 2'F-ANAs, also introduces unfavourable interactions between the 2'-fluorine of the modified riboses and neighboring O5' oxygens (2'F...O5' distances less than 3Å). This effect is supported by the upfield shift of the ¹⁹F signal of 2'F-RNA residues (Figure 5), and likely contributes to the lower stability of duplex **S** relative to **H**.

7.2.6. FC-H2'...O backbone interactions in other C2' fluorinated structures

Careful examination of others 2'F-modified structures in the literature reveals close C-H2'...O contacts. For example, in the crystallographic structures of a pure 2'F-RNA duplex (PDB:3P4A) and a chimeric 2'F-RNA/RNA duplex (PDB:3P4B), recently determined by Egli and co-workers²¹, H2'_{*i*}...O4'_{*i+1*} distances are around 2.6 Å (Tables 3, A7.7 and A7.8), and the C-H2'_{*i*}... O4'_{*i+1*} angles are around 140° (Figure 10).

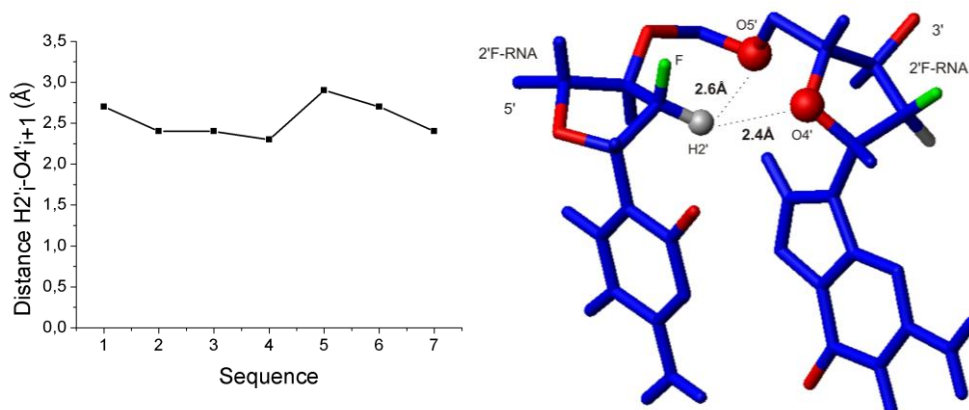


Figure 10: Left) Values of $H2'..O4'$ distances in 2'F-RNA:2'F-RNA crystallographic structure (PDB ID:3P4A). Right) View of short distances in 2'F-RNA/2'F-RNA steps in the crystallographic structure.

This geometry is fully consistent with hydrogen bond formation. The structures are pure A-form with all riboses and 2'F-ribose in the North conformation. Similarly, close $H2'..O4'_{i+1}$ distances and co-linear arrangement of $C-H2'..O4'_{i+1}$ is observed in the solution structure of a 2'F-ANA:RNA hybrid (Tables 3 and A7.10). In this case, the overall structure of the duplex is intermediate between A- and B-forms and 2'F-arabinoses adopt the East conformation²⁰. These contacts are not observed in the structure of 2'FANA:2'FANA duplexes, where the global structure is B-form and the arabinoses adopt the South conformation (see Table 3)¹⁹. Taking together all these observations, we conclude that sequential FC-H..O4' interactions are favoured in A-form duplexes (e.g., 2'F-RNA:2'F-RNA) or A-like hybrids (e.g. 2'F-ANA:RNA).

7.2.7. Are FC-H..O interactions real hydrogen bonds?

To address this question, quantum mechanical calculations in dinucleotide model systems that mimic the four possible step combinations that occur in **H** and **S**, *i.e.*, 2'F-RNA_i/2'F-RNA_{i+1}, 2'F-ANA_i/2'F-ANA_{i+1}, 2'F-RNA_i/2'F-ANA_{i+1} and 2'F-ANA_i/2'F-RNA_{i+1}, were carried out in Prof. Modesto Orozco's laboratory. 2'F-RNA_i/2'F-RNA_{i+1} and RNA_i/RNA_{i+1} dinucleotides from the high-resolution X-ray structures of a

pure 2'F-RNA duplex (PDB code 3P4A) and a RNA duplex (PDB code 1RXB), respectively^{21,24} were also included in the calculation.

In order to detect potential stabilizing interactions in these systems, the electron density of the optimized dinucleotide steps was analysed by means of the theory of “atoms in molecules”²⁵. Using this theoretical framework, the topology of the electron density was examined using the computer program EXTREME²⁶ to detect and characterize bond critical points (BCP). Here the analysis was restricted to bonds between substituent atoms at the C2' position of one nucleotide and the oxygen atom of the furanose ring in its 3'-neighbour nucleotide. Values of the electron density and the Laplacian of the electron density for the observed BCPs in each dinucleotide step are listed in Table 3. The presence of BCPs is a necessary condition to identify stabilizing noncovalent interactions. In the case of hydrogenbonds, the electron density in the BCP of a hydrogen-bond ranges from 0.002 to 0.011 atomic units (au)²⁷, and the Laplacian of the charge density is known to be positive and in the range of 0.014 to 0.139 au²⁷. Values shown in Table 3 are within this range, indicating the presence of a sequential FCH..O hydrogen bond, in all step types except 2'F-RNA_i/2'F-ANA_{i+1}.

Intra-molecular hydrogen bond strengths are difficult to estimate from the energetic AIM parameters. However, the values shown in Table 3 are in agreement with those reported for small molecular weight systems¹⁵, in which intermolecular C-H..O hydrogen bonds strength can be estimated with Espinosa relations²⁸. Consistent with the calculated AIM parameters in the different dinucleotides, FC-H..O hydrogen bonds in 2'F-RNA_i/2'F-RNA_{i+1} and 2'F-ANA_i/2'F-RNA_{i+1} steps are more stabilizing than C-H_i..O_{i+1} hydrogen bonds in RNA_i/RNA_{i+1} steps (Table 3 and A7.9). No hydrogen bond is detected in 2'F-RNA_i/2'F-ANA_{i+1} steps.

Table 3: Average values of $H2'/H2''_{i+1}..O4'$ distances (\AA), angles between $C-H2'/H2''_{i+1}..O4'$ in different structures, and topological properties of the bond critical points (BCP) detected between the $H2'/H2''_{i+1}..O4'$.

Step	Structure	Distance $H2'/H2''_{i+1}..O4'$	Angle C- $H2'/H2''_{i+1}..O4'$	ρ (au)	$\nabla^2\rho$ (au)
RNA _i /RNA _{i+1}	PureRNA (PDB: 1RXB)	2.5	144	0.005435	0.020660
2'F-RNA/2'F-RNA _{i+1}	Pure 2'F-RNA (PDB:3P4A)	2.6	140	0.010307	0.034542
2'F-RNA/2'F-RNA _{i+1}	H, S	2.9	110	0.007085	0.027376
2'F-ANA/2'F-ANA _{i+1}	H	2.8	160	0.004617	0.017697
2'F-ANA/2'F-RNA _{i+1}	S	2.5	147	0.007612	0.027338
2'F-RNA/2'F-ANA _{i+1}	S	4.4	73	Not detected	Not detected
2'F-ANA/2'F-ANA _{i+1}	2'F-ANA:RNA hybrid	2.9	146	n.c.	n.c.
2'F-RNA/2'F-RNA _{i+1}	Standard A- type	2.9	120	n.c.	n.c.
2'F-ANA/2'F-ANA _{i+1}	Standard B- type	4.7	107	n.c.	n.c.

n.c. Not calculated

7.2.8. 2'F effect in duplex stability: Backbone hydrogen bonds vs enhanced base stacking and base pairing

Increased Watson-Crick hydrogen bonding strength and enhanced staking interactions have been proposed as the main cause for 2'F-RNA stabilization relative to RNA²². Thermodynamic data from Egli and co-workers clearly show that the origin of 2'F-RNA stabilization is mainly enthalpic^{21,29}. NMR data from the same authors suggest that Watson-Crick hydrogen bonds in 2'F-RNA:2'F-RNA base pairs are stronger than in RNA:RNA base pairs, and they hypothesized that this is due to a polarization effect of the fluorine over the whole aromatic base, including the distant imino protons. This effect has not been studied in 2'F-ANA:2'F-ANA base pairs, however, our ¹H NMR melting experiments (shown in Figure 3), indicate that 2'F-RNA/2'F-RNA and 2'F-ANA/2'F-RNA base pairs have similar stability. Imino protons from both kinds of base pairs disappear at similar

temperatures. Also, **H** duplex is more stable than **S** duplex, even though there are more 2′F-RNA/2′F-RNA base pairs in the latter.

The different stabilities between duplexes containing different patterns of 2′F-ANA and 2′F-RNA substitution observed by Deleavey *et al.*²³ are difficult to explain only on the basis of base-pair strength or stacking interactions. In addition, a recent study by Pallan *et al.*²¹ shows that the increase in stability of chimeric 2′F-RNA/RNA duplex with alternated sequence, and pure 2′F-RNA duplexes is proportional to the number of 2′F-RNA in the sequence, although in the former there is no stacking between 2′F-RNAs and no 2′F-RNA:2′F-RNA base pairs. These observations are not consistent with the base polarization effects proposed by Patra *et al.*²². The structures of the chimeric 2′F-RNA:2′F-ANA duplexes reported here, as well as the quantum mechanical calculations on these and other structures, clearly indicate that the effect of polarization induced by fluorine in non-conventional FC-H...O hydrogen bond explains the enhanced stabilities observed in 2′F-RNA-containing duplexes. These interactions might also preorganize the single strand in a suitable conformation for duplex formation.

7.3. CONCLUSION

Chimeric oligonucleotides containing different combinations of 2′F-ANA and 2′F-RNA adopt solution structures intermediate between canonical A- and B- families of double helical structures. In continuous tracts fluoro-arabinoses adopt South/East-type sugar conformations, and fluoro-ribose C3′-endo conformations. However, when 2′F-RNA and 2′F-ANA are alternated in the sequence, fluoro-ribose adopt an unusual East-type conformation. In contrast to the case of 2′F-ANA:RNA hybrids, 2′F-ANA glycosidic angles adopt a low anti conformation in which 2′F...H8 pseudo-hydrogen bond formation is not possible. Alternation of 2′F-RNA and 2′F-ANA provokes a strong destabilization with respect to hybrids formed by continuous tracts. This destabilization is mainly due to changes in electrostatic interactions between 2′-substituents and the 3′-neighboring nucleotide. In tracts of continuous 2′F-ANA and 2′F-RNA sequences, the positive polarization of the

proton at 2'-position induced by the 2'F provokes favorable interactions with the sequential O4'. Quantum mechanical calculation indicate that the interaction FC-H2'/H2''...O4' is a nonconventional hydrogen bond.

BIBLIOGRAPHY

- (1) Kool, E. T. *Hydrogen Bonding, Base Stacking, And steric effects in DNA replication* **Annu. Rev. Biophys. Biomol. Struct.** (2001), 30, 1.
- (2) Šponer, J.; Leszczynski, J.; Hobza, P. *Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases* **Biopolymers** (2001), 61, 3.
- (3) Kool, E. T. *Preorganization of DNA: Design Principles for Improving Nucleic Acid Recognition by Synthetic Oligonucleotides* **Chem. Rev.** (1997), 97, 1473.
- (4) Ramachandran, G. N.; Sasisekharan, V. *Refinement of the structure of collagen* **Biochim. Biophys. Acta** (1965), 109, 314.
- (5) Krimm, S. *Hydrogen bonding of C - h * * * o = C in proteins* **Science** (1967), 158, 530.
- (6) Bond, C. S.; Zhang, Y.; Berriman, M.; Cunningham, M. L.; Fairlamb, A. H.; Hunter, W. N. *Crystal structure of Trypanosoma cruzi trypanothione reductase in complex with trypanothione, and the structure-based discovery of new natural product inhibitors* **Structure** (1999), 7, 81.
- (7) Bach, R. D.; Thorpe, C.; Dmitrenko, O. *C-H...Carboxylate Oxygen Hydrogen Bonding in Substrate Activation by Acyl-CoA Dehydrogenases: Synergy between the H-bonds* **J. Phys. Chem. B** (2002), 106, 4325.
- (8) Horowitz, S.; Trievel, R. C. *Carbon-Oxygen Hydrogen Bonding in Biological Structure and Function* **J. Biol. Chem.** (2012), 287, 41576.
- (9) Wahl, M. C.; Sundaralingam, M. *C-H...O hydrogen bonding in biology* **Trends Biochem. Sci.** (1997), 22, 97.
- (10) Benevides, J. M.; Thomas, G. J. *A solution structure for poly(rA)-poly(dT) with different furanose pucker and backbone geometry in rA and dT strands and intrastrand hydrogen bonding of adenine 8CH* **Biochemistry** (1988), 27, 3868.
- (11) Ghosh, A.; Bansal, M. *C-H...O hydrogen bonds in minor groove of A-tracts in DNA double helices* **J. Mol. Biol.** (1999), 294, 1149.
- (12) Quinn, J. R.; Zimmerman, S. C.; Del Bene, J. E.; Shavitt, I. *Does the A-T or G-C Base-Pair Possess Enhanced Stability? Quantifying the Effects of CH...O Interactions and Secondary Interactions on Base-Pair Stability Using a Phenomenological Analysis and ab Initio Calculations* **J. Am. Chem. Soc.** (2007), 129, 934.
- (13) Berger, I.; Egli, M.; Rich, A. *Inter-strand C-H...O hydrogen bonds stabilizing four-stranded intercalated molecules: stereoelectronic effects of O4' in cytosine-rich DNA* **Proc. Natl. Acad. Sci. U.S.A.** (1996), 93, 12116.
- (14) Berger, I.; Egli, M. *The Role of Backbone Oxygen Atoms in the Organization of Nucleic Acid Tertiary Structure: Zippers, Networks, Clamps, and C-H...O Hydrogen Bonds* **Chem. Eur. J.** (1997), 3, 1400.
- (15) Gu, Y.; Kar, T.; Scheiner, S. *Fundamental Properties of the CH...O Interaction: Is It a True Hydrogen Bond?* **J. Am. Chem. Soc.** (1999), 121, 9411.
- (16) Cubero, E.; Orozco, M.; Luque, F. J. *Electron density topological analysis of the C-H...O anti-hydrogen bond in the fluoroform-oxirane complex* **Chem. Phys. Lett.** (1999), 310, 445.
- (17) Khakshoor, O.; Wheeler, S. E.; Houk, K. N.; Kool, E. T. *Measurement and Theory of Hydrogen Bonding Contribution to Isosteric DNA Base Pairs* **J. Am. Chem. Soc.** (2012), 134, 3154.
- (18) Ikeda, H.; Fernandez, R.; Barchi, J. J.; Huang, X.; Marquez, V. E.; Wilk, A. *The effect of two antipodal fluorine-induced sugar puckers on the conformation and stability of the Dickerson-Drew dodecamer duplex [d(CGCGAATTCGCG)]2* **Nucleic Acids Res** (1998), 26, 2237.
- (19) Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; Gonzalez, C. *The solution structure of double helical arabino nucleic acids (ANA and 2'F-ANA): effect of arabinoses in duplex-hairpin interconversion* **Nucleic Acids Res** (2012), 40, 9329.

- (20) Watts, J. K.; Martín-Pintado, N.; Gómez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; González, C.; Damha, M. J. *Differential stability of 2'-F-ANA•RNA and ANA•RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility* **Nucleic Acids Res** (2010), 38, 2498.
- (21) Pallan, P. S.; Greene, E. M.; Jicman, P. A.; Pandey, R. K.; Manoharan, M.; Rozners, E.; Egli, M. *Unexpected origins of the enhanced pairing affinity of 2'-fluoro-modified RNA* **Nucleic Acids Res** (2011), 39, 3482.
- (22) Patra, A.; Paolillo, M.; Charisse, K.; Manoharan, M.; Rozners, E.; Egli, M. *2'-Fluoro RNA Shows Increased Watson–Crick H-Bonding Strength and Stacking Relative to RNA: Evidence from NMR and Thermodynamic Data* **Angew. Chem. Int. Ed.** (2012), 51, 11863.
- (23) Deleavey, G. F.; Watts, J. K.; Alain, T.; Robert, F.; Kalota, A.; Aishwarya, V.; Pelletier, J.; Gewirtz, A. M.; Sonenberg, N.; Damha, M. J. *Synergistic effects between analogs of DNA and RNA improve the potency of siRNA-mediated gene silencing* **Nucleic Acids Res** (2010), 38, 4547.
- (24) Portmann, S.; Usman, N.; Egli, M. *The Crystal Structure of r(CCCCGGGG) in Two Distinct Lattices* **Biochemistry** (1995), 34, 7569.
- (25) Bader, R.; Molecules, A. I. *A quantum theory*, Clarendon: Oxford, UK, (1990).
- (26) Biegler-König, F. *Calculation of atomic integration data* **J. Comput. Phys.** (2000), 21, 1040.
- (27) Cubero, E.; Orozco, M.; Hobza, P.; Luque, F. J. *Hydrogen Bond versus Anti-Hydrogen Bond: A Comparative Analysis Based on the Electron Density Topology* **J. Phys. Chem. A** (1999), 103, 6394.
- (28) Espinosa, E.; Alkorta, I.; Rozas, I.; Elguero, J.; Molins, E. *About the evaluation of the local kinetic, potential and total energy densities in closed-shell interactions* **Chem. Phys. Lett.** (2001), 336, 457.
- (29) Manoharan, M.; Akinc, A.; Pandey, R. K.; Qin, J.; Hadwiger, P.; John, M.; Mills, K.; Charisse, K.; Maier, M. A.; Nechev, L.; Greene, E. M.; Pallan, P. S.; Rozners, E.; Rajeev, K. G.; Egli, M. *Unique Gene-Silencing and Structural Properties of 2'-Fluoro-Modified siRNAs* **Angew. Chem. Int. Ed.** (2011), 123, 2332.

Chapter 8

Effect of incorporation of 2'-deoxy-2',2'-difluorodeoxycytidine (gemcitabine) on the structure and stability of DNA:RNA hybrid duplexes

RESUMEN

En este capítulo se estudia el efecto de insertar un nucleótido de 2'-deoxy-2'2'-difluorodeoxycytidina (dFdC) en la cadena de DNA de un dúplex híbrido DNA:RNA. Aunque esta modificación no afecta significativamente la estabilidad del dúplex híbrido, la estructura de dicho dúplex muestra que el nucleótido difluorado en C2' provoca distorsiones que afectan varios pares de bases cerca del lugar de la modificación. DFdC tiende a adoptar una conformación norte en dúplex de DNA o RNA. Sin embargo, dFdC adopta una conformación sur en el contexto de dúplex híbridos. Esta conformación del azúcar favorece la formación de una interacción atractiva 2'F..H8, que compensa parcialmente la desestabilización provocada por las distorsiones en los pares de bases. Estas interacciones comparten algunas características de los pseudo-enlaces de hidrógeno descritos en el capítulo 6.

ABSTRACT

In this chapter we study the effect of a single 2'-deoxy-2',2'-difluorodeoxycytidine (dFdC) insertion in the DNA strand of a DNA:RNA hybrid duplex. Although this modification does not significantly affect the stability of the hybrid duplex, its solution structure shows that the C2'-difluorinated nucleotide provokes distortions in the duplex that affect several base pairs near the site of the substitution. DFdC tends to adopt North type sugar conformation when isolated or inserted in pure DNA or RNA duplex. However, dFdC adopts a South sugar conformation in the context of DNA:RNA hybrids. This sugar pucker favors the formation of a 2'F..H8 attractive interaction, which compensates the destabilizing effect of base pair distortions. These interactions share some features with pseudohydrogen bonds described previously in Chapter 6.

8.1. INTRODUCTION

As shown in the previous chapters, C2'-fluorine modified nucleotides alters the stability of duplexes through several kind of interactions^{1,2}. Fluorine high electronegativity changes the charge distribution and favors 2'F..H8 or FC-H2'/H2''..O electrostatic interactions depending on the structural context³. Interestingly, optimization of these interactions is totally dependent on the structure and affects duplex stability. One of the clearest examples is observed with 2'F-ANA nucleotides; while 2'F-ANA:2'F-ANA homoduplex adopts a B-form structure and disrupts inter-nucleotide 2'F..H8 and FC-H2''..O interactions², 2'F-ANA:RNA hybrid duplex adopts an intermediate structure and favors both types of attractive interactions¹.

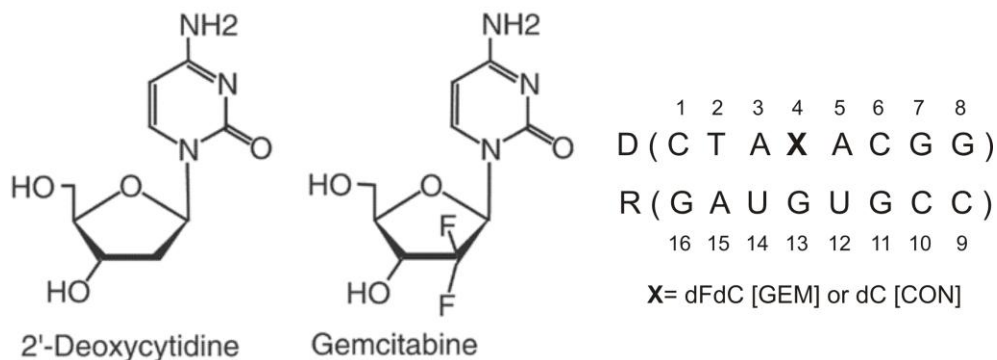


Figure 1: Left) Schematic structures of DNA and dFdC. Right) Numbering scheme of the 8-mer hybrid duplexes used for this study. X indicates DNA or dFdC substitutions. The corresponding hybrids are named **[CON]** and **[GEM]** respectively.

In this chapter we explore the effect of a double fluorine substitution in the C2' position in a hybrid duplex. 2'-deoxy-2',2'-difluorodeoxycytidine (dFdC), also named gemcitabine, which is a modified nucleotide currently in use in anticancer treatments (Figure 1). dFdC has shown efficiency against a wide range of solid tumors⁴⁻⁸ and normally is used in combination chemotherapy treatments^{7,9,10}. Nowadays, dFdC's mechanism of action is not totally clear, although the most accepted hypothesis suggests that the triphosphate form of the gemcitabine (dFdCTP) acts as a competitive substrate of deoxycytidine and is incorporated into

DNA during replication. After the incorporation of dFdC, only one extra nucleotide is added before DNA polymerase is unable to continue replicating¹¹. This interference in the chain elongation of DNA causes the cell death by apoptosis¹².

Previous studies have shown that incorporation of a single dFdC in DNA:DNA and RNA:RNA duplexes destabilizes the structure^{13,14}. Additionally, Konerding *et al.* determined that, dFdC alters the electrostatic charges around the modification in an Okazaki fragment, provoking only minor local changes in the DNA:DNA homoduplex region¹⁴. Surprisingly, the most significant distortions were observed in the DNA/RNA junction of the Okazaki fragment, located three nucleotides apart from the modification.

To gain further insights on the perturbation provoked by this modification in DNA:RNA hybrids, we have carried out the three dimensional structural determination of a modified hybrid octamer, with one dFdC in the DNA strand **[GEM]** (Figure 1). The unmodified DNA:RNA hybrid duplex **[CON]** was also solved in the same experimental conditions for structural comparison. Our results show that one insertion of dFdC provokes clear distortions in local and adjacent base pairs and favors 2'F..H8 interactions to stabilize the intermediate structure between A and B-form.

8.2. RESULTS

8.2.1. Effect of dFdC modification in hybrid duplex stability

Duplex formation and melting of both **[GEM]** and **[CON]** were monitored by CD and NMR. T_m values from CD meltings are shown in Table 1 and reflect that both duplexes have very similar stabilities ($\Delta T_m=1^\circ\text{C}$). Comparison of **[GEM]** and its corresponding control duplex CD spectra shows a slight broader maxima at 265 nm in the duplex control. This effect indicates that **[GEM]** is more A-like than **[CON]** in the same experimental conditions (Figure 2)¹⁵.

Table 1: Melting temperatures of **[GEM]** duplex and its control **[CON]** (buffer conditions 25 mM sodium phosphate, 100 mM NaCl, pH 7).

Name	Sequence (5'-3') ^a	$T_m(^{\circ}\text{C})$
[GEM]	5'-CTA(dFdC)ACGG-3'	30.4
[CON]	5'-CTACACGG-3'	31.0

^aThe DNA strand is represented in the table. The complementary strand is RNA.

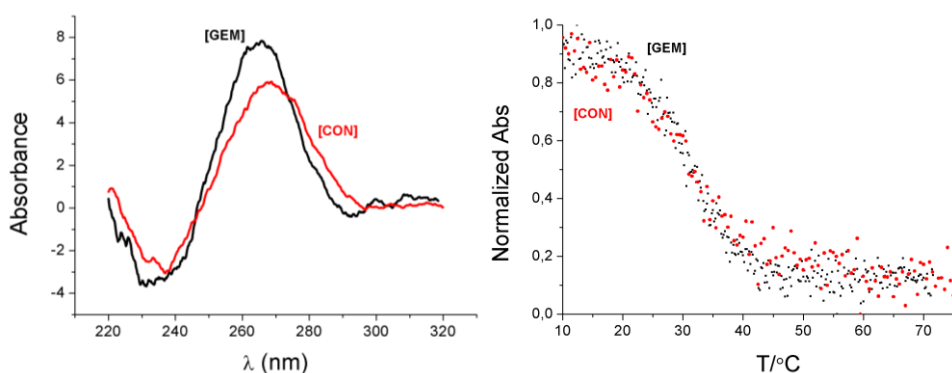


Figure 2: **[GEM]** and **[CON]** CD spectra at 4°C (left) and CD-monitored melting (right). **[GEM]** modified duplex is shown in black and **[CON]** hybrid duplex in red. Duplex concentration 4 μM (25 mM sodium phosphate, 100 mM NaCl, pH=7).

NMR imino region is also consistent with duplex formation in both cases. Melting experiments carried out on the NMR samples confirmed that both duplexes have very similar thermal denaturation profile and T_m in the same experimental conditions (Figure 3). The imino region of both samples shows NOE cross-peak patterns of typical Watson-Crick base pairs (Figure 4).

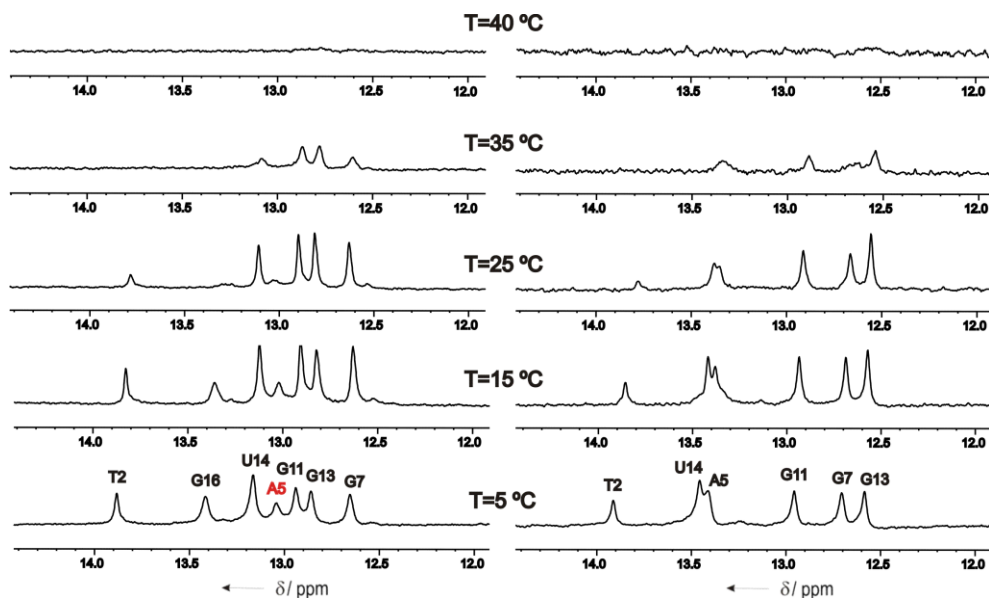


Figure 3: Left) ^1H spectra of **[GEM]** at different temperatures. Right) ^1H spectra of **[CON]** at different temperatures. Oligonucleotide concentration 0.6 mM (25 mM sodium phosphate, 100mM NaCl, pH 7, $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1). Imino of A5 in **[GEM]** is labeled in red.

8.2.2. NMR assignments

Due the good spectral resolution in both samples, sequential assignments were obtained following standard methods for right-handed, double-stranded nucleic acids using DQF-COSY, TOCSY and 2D NOESY spectra (Figure 4). The assignments pathways could be followed in the base- $\text{H1}'$ (Figure 5) and in the base- $\text{H2}'/\text{H2}''$ regions, except for dFdC modification. Complete assignments could be carried out with the exception of $\text{H5}'/\text{H5}''$ and some overlapped $\text{H3}'$ and $\text{H4}'$ signals in the RNA strand.

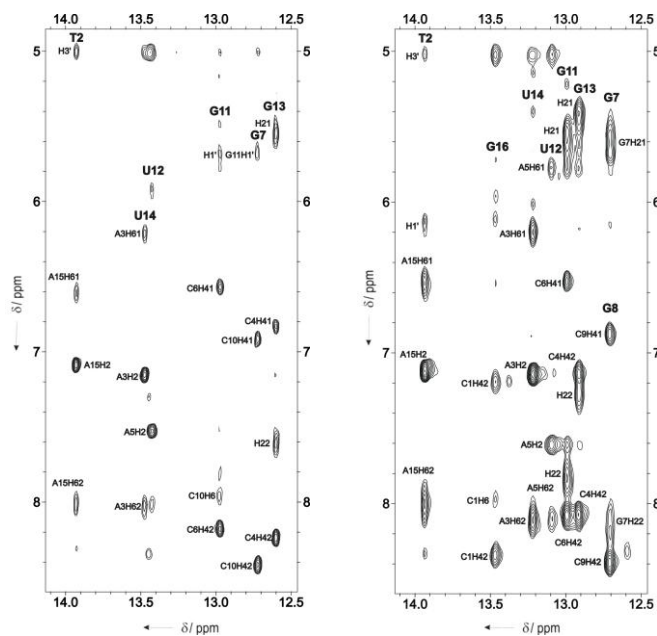


Figure 4:Imino region of the NOESY spectra in H₂O (τ_m =100ms) of **[CON]** (left), and **[GEM]** (right). Buffer conditions 100 mM NaCl, 25 mM sodium phosphate, T=10°C, pH 7.

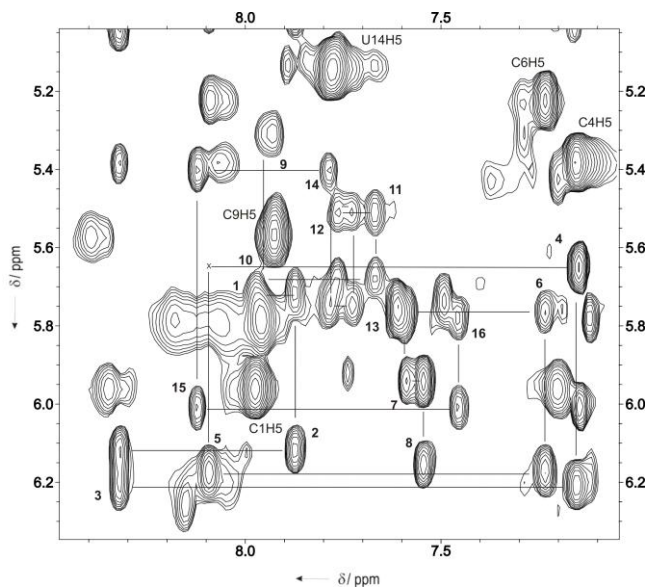


Figure 5: Base-H1' region of the NOESY spectra in H₂O ($\tau_m=100$ ms) of **[GEM]**. Buffer conditions were 100 mM NaCl, 25 mM sodium phosphate, T=25 °C, pH 7.

Exchangeable protons were assigned with the NOESY spectra recorded in H₂O (Figure 4). Most of the labile protons were assigned following standard methods, except some amino resonances of guanines that were not detected. The cross-peak patterns observed for the exchangeable protons indicate that all bases are forming Watson-Crick pairs throughout the duplex. Chemical shifts of some exchangeable protons change in **[GEM]** and **[CON]** duplexes, especially in the modified diFC4:G13 (0.3 ppm) and the adjacent A3:U14 (0.2 ppm) and A5:U12 (0.4 ppm) base pairs, what suggests structural differences provoked by the gemcitabine. Interestingly, A5:U12 imino signal is also broader and seems to disappear at lower temperature than the rest of **[GEM]** signals, which might reflect that the dFdC also affects the strength or the dynamics of 3'-sequential base pair (Figure 3, A5 is labeled in red). The full assignments are given in Table A8.1 and A8.2.

The **[GEM]** ¹H spectrum decoupled to ¹⁹F shows two broad doublets with a coupling constant of 245 Hz between germinal fluorines. Broadening of the signals at low temperatures, along with the splitting, impeded the acquisition of 2D heteronuclear HOESY experiments.

8.2.3. Experimental constraints and structure calculations

Quantitative distance constraints were obtained from NOESY experiments by using complete relaxation matrix analysis¹⁶. Around 140 structurally relevant experimental distance constraints were obtained (Table 2). Some structural information can be determined from these distances. Sequential H1'-base NOEs in the RNA strand are consistent with a standard A-form duplex. However, many interproton distances in DNA strand are in disagreement with a pure A conformation. For example, the strong sequential H2''-H6/H8 NOEs indicate significant B-like conformation in the DNA strand of both duplexes. Experimental evidences of sugar conformation of dFdC modification are difficult to obtain owing to the lack of H2' and H2'' assignments. However, the low intra-residual intensity of H3'-H6 NOE is characteristic of the South conformation of the sugar.

Table S3: NMR restraints and structural statistics.

Experimental ^1H - ^1H constraints	[GEM]	[CON]
Total number	145	137
Intra-residue	83	76
Sequential	34	46
Inter-strand	28	15
RMSD (Å)⁺		
all heavy atoms	0.6 ± 0.1	0.6 ± 0.2
backbone	0.6 ± 0.2	0.6 ± 0.2
all bases	0.6 ± 0.1	0.5 ± 0.1
sum of violations (Å) (average and range)	2.5 (2.2-2.9)	7.2 (6.8-7.7)
max. violation (average and range)	0.3 (0.2-0.3)	0.3 (0.3-0.4)
NOE energy kcal/mol (average and range)	9.2 (12-8.2)	30 (30-29)

⁺ All except terminal residue

In addition to the NOE-derived information, a qualitative analysis of the J-coupling constants obtained from DQF-COSY spectra was carried out. $J_{1'2'}$ are undetectable in most of the riboses of the RNA strand of both duplexes, which strongly suggests that riboses are in a pure North conformation. In the DNA strand, however, $J_{1'2'}$ and $J_{3'2'}$ point towards South/East conformations. Accordingly, torsion angle constraints were set for the dihedral angles of the sugars. The riboses of the RNA were constrained to North conformation. In DNA strand, sugar dihedrals were constrained to avoid North conformation and allow East/South regions. Backbone and dFdC dihedral angles were not constrained in the calculations. No scalar couplings were detected between fluorine atoms and H6/H8 protons (chapter 5 and 6)^{2,17}.

All the distance and torsion angle constraints were used to calculate the structure of both hybrid duplexes by restrained molecular dynamics as described in Annex I. The 10 final structures of both hybrids are displayed in Figure 6. As can be seen in the figure and in Table 2, both duplexes are well defined with a RMSD of 0.6 Å (excluding terminal residues). The final AMBER energies and NOE terms are reasonably low in all structures, with no distance constraint violation greater than 0.3 Å.

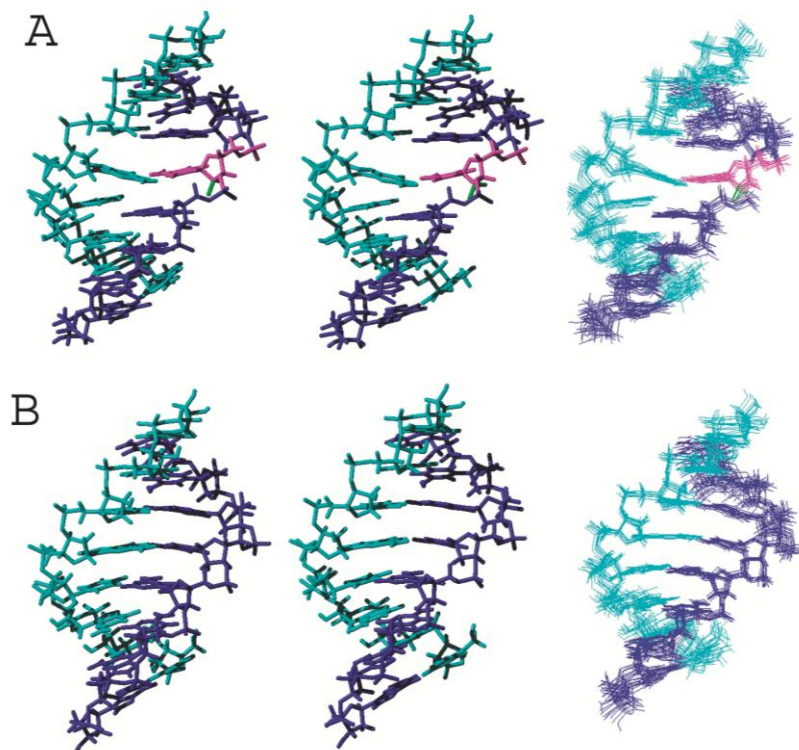


Figure 6: Superposition of 10 refined structures (right) and the stereoviews of the average structure (left). A) Structures of **[GEM]** and B) structures of **[CON]**. Blue indicates DNA strand and cyan RNA strand. DFdC4 is in magenta and fluorines in green.

8.2.4. Comparison of the structures

The overall shapes of **[GEM]** and **[CON]** is intermediate between A and B-form duplexes, being both of them closer to A-form. The RMSD between the average structure of **[GEM]** and an A-form duplex is 1.7 Å and 3.5 Å with a B-form duplex. The RMSD between **[GEM]** and its control **[CON]** is 1.3 Å for heavy atoms and non-terminal residues.

In both duplexes, helical parameters present intermediate values between A and B-form, i.e. rise values are around 3.4 Å typical of B-form and twist values around 32°, characteristic of A-form duplexes. The minor groove width is also intermediate between North and South conformations (Figure 7). Interestingly, **[GEM]** minor

groove is slightly more A-type than **[CON]**, which is in agreement with CD spectra differences between both samples (Figure 2).

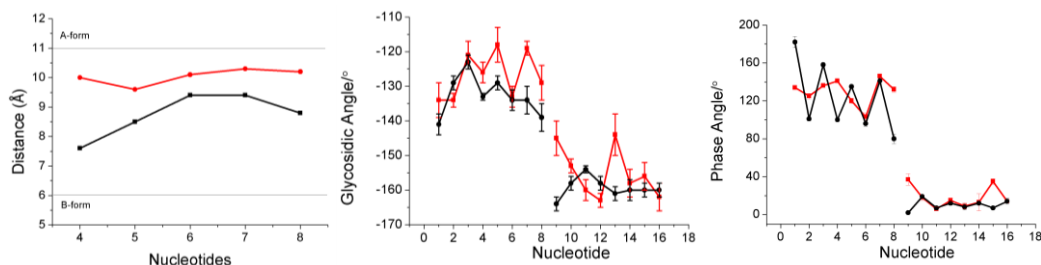


Figure 7: Values of minor groove width (left), glycosidic angle (middle) and phase angle (right) along the sequence of **[GEM]** and **[CON]**. **[GEM]** is in red and **[CON]** is in black. Standard A-form and B-form are shown in the minor groove for comparison.

The geometry of the RNA strand is similar in both hybrids, with all riboses in North conformation (C3'-endo) and glycosidic angles around -160° . Interestingly, the only exception is G13 in **[GEM]** duplex, with a smaller glycosidic angle around -143° . In the DNA strand of both duplexes phase angles are around -100° to 150° (South conformation) and glycosidic angles between -120° and 130° . The dFdC sugar adopts a South/East conformation (pseudorotation phase angle around 141°) as indicated by the weak NOE H3'-H6. A smaller glycosidic angle is also observed in dFdC (around 118°) compared with its control (134°). The rest of backbone parameters are very similar in both duplexes. The only exception is zeta dihedral angle with higher values in **[GEM]** duplex. Complete tables of geometrical parameters are shown in Tables A8.3 and A8.4.

Significant distortions are observed in dFdC4:G13 and surrounding base pairs (Figure 8). Such distortions are reflected in the helical parameters of these base pairs. Thus, base-axis, base-base and inter-base parameters show drastic variations in the modified dFdC4-G13 pair and its 3' (A3-U14) or 5' (A5-U12) adjacent base pairs. The most important changes are observed in Inclination, Tip, Buckle and Tilt values (Tables A8.5-A8.8). These base pairs distortions are in full agreement with imino changes observed in **[GEM]** (Figure 3).

8.3. DISCUSSION

8.3.1. Structural comparison between [GEM] and [CON]

Despite the structural similarities between [GEM] and its control hybrid duplex, significant distortions have been found in the modified duplex. The most striking difference is the distortion in dFdC4:G13 modified base pair compared with [CON]. An observable kink in the base pair provokes important changes in buckle parameters; difference between [GEM] and [CON] duplexes is around 42° (Figure 8). Moreover, this perturbation strongly affects the 3'-sequential A5:U12 pair disrupting the planarity between bases, which might demises the strength of this base pair's interaction (Figure 8). This structural effect totally agrees with the broadening and lower temperature disappearance of A5:U12 imino signal compared with the rest of [GEM] imino signals. This data suggests weaker base pairs interactions or different dynamics of the residues involved in the interaction(Figure 3).

Computational and crystallographic studies have shown the single dFdC nucleotide tends to adopt a North sugar pucker¹⁸. A structural study of a single insertion of gemcitabine in DNA homoduplex shows that dFdC also adopts a North conformation in this context¹⁴. Such sugar pucker does not provoke any distortion or destabilizing interaction. Interestingly, a more exhaustive conformational analysis on 2'-deoxy-2'2'-difluoro-4'-dihydro-4'-thionucleosides has proved that weak interactions, such as dipole-dipole or electrostatic attractions, are essential to determine the overall conformation of these sugar rings¹⁹. Thus, NMR, high-level quantum calculations and molecular dynamics show that weak interactions, as 2'F..H6 electrostatic attractions, favor the South conformation of the sugar. In contrast, the packing forces (OH..O=C hydrogen bonds or 2'F..C=O interactions) in the solid state compensate the energetic disadvantages of the less stable North conformation and drive the thionucleoside to crystallize in North conformation. Thus, these results suggest that C2'-difluorinated nucleosides might adopt different conformations depending on their chemical and structural context to optimize favorable electrostatic interactions.

In **[GEM]**, dFdC4 nucleotide adopts a southern (phase angle around 141°) conformation. This South ($C2'$ -endo) sugar pucker of dFdC in the DNA strand optimizes the geometry to form $2'F..H8$ inter-nucleotide pseudohydrogen bonds in pyrimidine-purine steps (Figure 8C)¹. Thus, one of the fluorines is close to its own H6 of the base ($d=2.4$ Å), while the other might form an inter-residue pseudohydrogen bond ($d=2.4$ Å). Moreover, a structure calculation was carried out blocking the dFdC sugar in a North conformation. The resulting model structure shows very short $2'F..O$ contacts (2.9 Å) which might provoke steric clashes and destabilizing interactions (Figure 8). Thus, these destabilizing interactions in North conformation, together with the possibility to form stabilizing pseudohydrogen bonds in South conformation might promote the southern dFdC sugar puckering DNA:RNA hybrids.

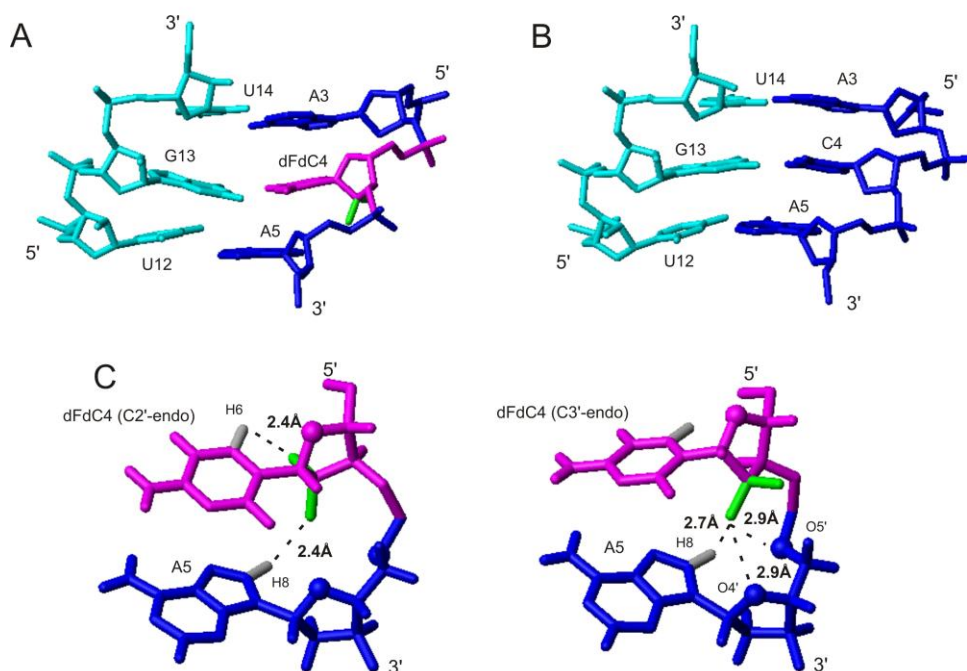


Figure 8: Average structure of A3:U14, C4:G13 and A5:U12 base pairs in A) **[GEM]** and B) **[CON]**. C) $2'F..H8/H6$ and $2'F..O$ distances comparison between dFdC4 in South (left) and North (right) conformation. RNA is shown in cyan, DNA in blue and dFdC in magenta. Fluorine atoms are in green and H6 and H8 in gray.

These structural changes provoked by dFdC also explain the similar stability of both duplexes. While base pairs distortions (dFdC4:G13 and A5:U12) destabilize the duplex, stabilizing 2'F..H8 interactions might compensate this effect. These changes in the sugar pucker confirm the ability of C2'-difluorine modified nucleotides to adopt the most favorable conformation to optimize stabilizing interactions in each particular structural context.

8.3.2. 2'F..H8 stabilizing interactions

The importance of non-conventional 2'F..H8 interactions has been described previously in Chapter 5 and 6. While experimental evidence for inter-nucleotide pseudohydrogen bonds arise from strong 2'F-H8 HOE cross-peaks and fluorine chemical shifts variations *versus* temperature (chapter 6), intra-nucleotide interactions has been demonstrated by heteronuclear J-couplings between 2'-fluorine and purine H8 of the same nucleotide. Moreover, the NMR structures strongly suggest the presence of these stabilizing interactions.

In the **[GEM]** duplex, the lack of heteronuclear J-couplings between 2'F and pyrimidine H6 suggests an absence of intra-nucleotide interactions. On the other hand, the broadening of fluorine signals at low temperatures impeded the acquisition of HOESY experiments but the inter-nucleotide pseudohydrogen bond 2'F..H8 in pyrimidine-purine steps seems to be structurally optimized ($d=2.4$ Å). It is important to mention that this interaction might be more intense in 2'F-ANA/2'F-ANA steps than in 2'F-ANA/DNA ones, since 2'F-ANA nucleotides polarize the base and acidify the H8 of the purines. Thus, **[GEM]** stabilization by 2'F..H8 interactions might be lower than in 2'F-ANA:RNA duplex reported previously (chapter 6).

8.4. CONCLUSION

We have determined the structural changes provoked by an insertion of a dFdC in a DNA:RNA hybrid duplex. **[GEM]** and **[CON]** have a very similar intermediate

structure with local distortions in the modified base pair (dFdC4:G13) and its adjacent bases. DFdC provokes a kink in its own base pair that affects the stability of 3'-adjacent base pair. DFdC adopts a South conformation that favors 2'F..H8 interactions, similar as those reported previously (chapter 6). This is interesting since dFdC adopts a North conformation when isolated or inserted in a DNA:DNA duplex. Further analysis of dFdC4 in a North conformation revealed that destabilizing 2'F..O interactions likely impede the formation of this sugar pucker in intermediate structures. Therefore, as other C2'-fluorine modified nucleotides, dFdC is able to adopt the most favorable conformation to optimize stabilizing interactions. These stabilizing interactions might balance the destabilization provoked by the base pairs distortion thus account for the thermal similar stability of [GEM] and [CON].

BIBLIOGRAPHY

- (1) Watts, J. K.; Martín-Pintado, N.; Gómez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; González, C.; Damha, M. J. *Differential stability of 2'F-ANA•RNA and ANA•RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility* **Nucleic Acids Res** (2010), 38, 2498.
- (2) Martín-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; Gonzalez, C. *The solution structure of double helical arabinonucleic acids (ANA and 2'F-ANA): effect of arabinoses in duplex-hairpin interconversion* **Nucleic Acids Res** (2012), 40, 9329.
- (3) Martín-Pintado, N.; Yahyaee-Anzahaee, M.; Deleavey, G. F.; Portella, G.; Orozco, M.; Damha, M. J.; Gonzalez, C. *Dramatic effect of furanose c2' substitution on structure and stability: directing the folding of the human telomeric quadruplex with a single fluorine atom* **J. Am. Chem. Soc.** (2013), 135, 5344.
- (4) Carmichael, J.; Walling, J. *Phase II activity of gemcitabine in advanced breast cancer* **Semin. Oncol.** (1996), 23, 77.
- (5) Carmichael, J.; Fink, U.; Russell, R. C. G.; Spittle, M. F.; Harris, A. L.; Spiessi, G.; Blatter, J. *Phase II study of gemcitabine in patients with advanced pancreatic cancer* **Br. J. Cancer** (1996), 73, 101.
- (6) Hoang, T.; Kim, K. M.; Jaslowski, A.; Koch, P.; Beatty, P.; McGovern, J.; Quisumbing, M.; Shapiro, G.; Witte, R.; Schiller, J. H. *Phase II study of second-line gemcitabine in sensitive or refractory small cell lung cancer* **Lung Cancer** (2003), 42, 97.
- (7) Yardley, D. A. *Gemcitabine plus paclitaxel in breast cancer* **Semin. Oncol.** (2005), 32, S14.
- (8) Hertel, L. W.; Boder, G. B.; Kroin, J. S.; Rinzel, S. M.; Poore, G. A.; Todd, G. C.; Grindey, G. B. *Evaluation of the antitumor-activity of gemcitabine (2',2'-difluoro-2'-deoxycytidine)* **Cancer Res.** (1990), 50, 4417.
- (9) Ozols, R. F. *Gemcitabine and carboplatin in second-line ovarian cancer* **Semin. Oncol.** (2005), 32, S4.
- (10) Peters, G. J.; van der Wilt, C. L.; van Moorsel, C. J. A.; Kroep, J. R.; Bergman, A. M.; Ackland, S. P. *Basis for effective combination cancer chemotherapy with antimetabolites* **Pharmacol. Ther.** (2000), 87, 227.
- (11) Huang, P.; Chubb, S.; Hertel, L. W.; Grindey, G. B.; Plunkett, W. *ACTION OF 2',2'-DIFLUORODEOXYCYTIDINE ON DNA-SYNTHESIS* **Cancer Res.** (1991), 51, 6110.
- (12) Moysan, E.; Bastiat, G.; Benoit, J. P. *Gemcitabine versus Modified Gemcitabine: A Review of Several Promising Chemical Modifications* **Mol. Pharm.** (2013), 10, 430.
- (13) Sierant, M.; Sobczak, M.; Janicka, M.; Paduszynska, A.; Piotrkowska, D. *Biological and physicochemical characterization of siRNAs modified with 2',2'-difluoro-2'-deoxycytidine (gemcitabine)* **New J. Chem.** (2010), 34, 918.
- (14) Konerding, D.; James, T. L.; Trump, E.; Soto, A. M.; Marky, L. A.; Gmeiner, W. H. *NMR structure of a gemcitabine-substitute model Okazaki fragment* **Biochemistry** (2002), 41, 839.

- (15) Kypr, J.; Kejnovska, I.; Renciuik, D.; Vorlickova, M. *Circular dichroism and conformational polymorphism of DNA* **Nucleic Acids Res** (2009), 37, 1713.
- (16) Borgias, B. A.; James, T. L. *MARDIGRAS-A procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure* **J. Magn. Reson.** (1990), 87, 475.
- (17) Anzahaee, M. Y.; Watts, J. K.; Alla, N. R.; Nicholson, A. W.; Damha, M. J. *Energetically Important C-H center dot center dot center dot F-C Pseudohydrogen Bonding in Water: Evidence and Application to Rational Design of Oligonucleotides with High Binding Affinity* **J. Am. Chem. Soc.** (2011), 133, 728.
- (18) Hausheer, F. H.; Jones, N. D.; Seetharamulu, P.; Singh, U. C.; Deeter, J. B.; Hertel, L. W.; Kroin, J. S. *Ab initio quantum mechanical and X-ray crystallographic studies of gemcitabine and 2'-deoxy cytosine* **Comput. Chem.** (1996), 20, 459.
- (19) Zheng, F.; Fu, L.; Wang, R. X.; Qing, F. L. *Synthesis and conformational analysis of D-2 '-deoxy-2 ',2 '-difluoro-4 '-dihydro-4 '-thionucleosides* **Org. Biomol. Chem.** (2010), 8, 163.

Chapter 9

Effect of furanose C2'-substitution on structure and stability
of G-quadruplex

RESUMEN

La secuencia del telómero humano puede adoptar diferentes conformaciones en disolución. En concreto, la secuencia d(TAGGGTTAGGGT) adopta dos tipos de cuádruplex bimoleculares en equilibrio, uno paralelo y otro antiparalelo. En este capítulo hemos determinado que las modificaciones de ANA, 2'F-ANA y RNA estabilizan el cuádruplex paralelo, siendo 2'F-ANA la sustitución más estabilizante, lo que ha permitido la determinación de su estructura por RMN. El análisis de esta estructura nos ha permitido determinar que la estabilización de este tipo de cuádruplex se debe, fundamentalmente, a las interacciones 2'F..H8 y FC-H2''..H8 secuenciales. Sorprendentemente, los nucleótidos de 2'F-RNA desestabilizan drásticamente la formación del cuádruplex debido a interacciones repulsivas entre los grupos fosfato y el 2'F.

ABSTRACT

Human telomeric DNA quadruplex can adopt different conformations in solution. We have found that arabinose (ANA), 2'F-arabinose (2'F-ANA), and ribose (RNA) substitutions stabilize the propeller parallel G-quadruplex form over competing conformers, allowing NMR structural determination of this particularly significant nucleic acid structure. 2'F-arabinose substitution provides the greatest stabilization as a result of electrostatic (FC-H2''..O4') and pseudohydrogen bonding (2'F..H8) stabilizing interactions. In contrast, 2'F-RNA substitution provokes a dramatic destabilization of the quadruplex structure due to unfavorable electrostatic repulsion between phosphate and the 2'F.

9.1. INTRODUCTION

G-quadruplexes are nucleic acid structures occurring in human telomers and oncogene-promoter regions (see chapter 2, section 2.2.3.5.), and have garnered considerable attention due to their involvement in telomere maintenance and gene regulation¹⁻³. They have also emerged as novel building blocks for supramolecular assemblies⁴, DNA-based nanodevices⁵⁻⁹, potential antiviral agents, and as a new class of molecular targets for anticancer drugs^{10,11}. G-quadruplexes exhibit significant structural diversity^{3,12,13}. High resolution structural studies on human telomeric DNA have mainly focused on short DNA sequences able to form individual quadruplexes. In many cases, these oligonucleotide sequences can adopt and interconvert between different conformations, obscuring structural analysis. Among the observed telomeric G-quadruplex topologies, it has been suggested that the simple folds of the parallel “propeller” and “hybrid” (3+1)^{14,15} structural models provide more adequate scaffolds for the compact-stacking multi-quadruplex structure of human telomeric DNA than other antiparallel models^{12,13}. Interestingly, human telomeric RNA (TERRA) which is known to play a crucial role in telomere maintenance *in vivo*^{16,17} always adopts propeller-like parallel structures^{18,19}.

The tendency of TERRA and other RNA quadruplexes to form parallel structures is related to the strong preference of guanine ribonucleosides to adopt the *anti* glycosidic bond conformation²⁰. The *syn-anti* preference of dG (DNA) vs rG (RNA) has been used to modulate G-quadruplex folding, and different nucleoside modifications have been proposed following this strategy²⁰⁻²⁶. In particular, 2'-Deoxy-2'-fluoro-arabinonucleosides (e.g., 2'F-araG) favor the anti-orientation in order to avoid steric interaction between guanine and the top-face fluorine of the arabinose sugar^{22,27}. The 2'F-arabinose modification is especially interesting since it can modulate the structure and stability of quadruplexes^{22,23}, and unlike ribose, it does not disrupt the conformation of the furanose ring²⁸⁻³⁰. Thus, replacement of dGs adopting the *anti* glycosidic bond conformation with 2'F-araG's stabilize the thrombin binding aptamer²² and the “hybrid (3+1)” quadruplex²³ up to +3.4°C per 2'F-araG modification.

To expand upon these findings and to provide the structural basis of the effect of C2' modifications in G-quadruplexes, we have studied the impact of several substitutions, 2'F-araG (**AFtel**), 2'F-rG (**RFtel**), araG (**Atel**) and rG (**Rtel**), in the human telomeric quadruplex (Figure 1). In this study we have focused on the two-repeat telomeric sequence, d(TAGGGTTAGGGT) (**Dtel**), which offers a convenient model for high resolution studies, either by NMR³¹ or crystallography¹⁴. In solution this sequence exists as a mixture of two interconverting forms: a parallel "propeller"-like fold and an antiparallel fold³¹. We have now found that a single dG to 2'F-araG modification at the 9th position provides the propeller-like conformation exclusively, allowing for detailed structural analysis. Our rationale for substitution at G9 d(TAGGGTTAGGGT) stems from the observation that in antiparallel quadruplexes the guanine at the 3'-side of the first TTA loop tends to adopt a *syn* glycosidic bond conformation¹³, which is disfavored in 2'F-arabino nucleosides. The nucleotide structures and oligonucleotide sequences studied are shown in Table 1 and Figure 1. Modified nucleotides used in this chapter are guanosines. They are named as dG for DNA, rG for RNA, 2'F-araG for 2'F-ANA, araG for ANA and 2'F-rG for 2'F-RNA.

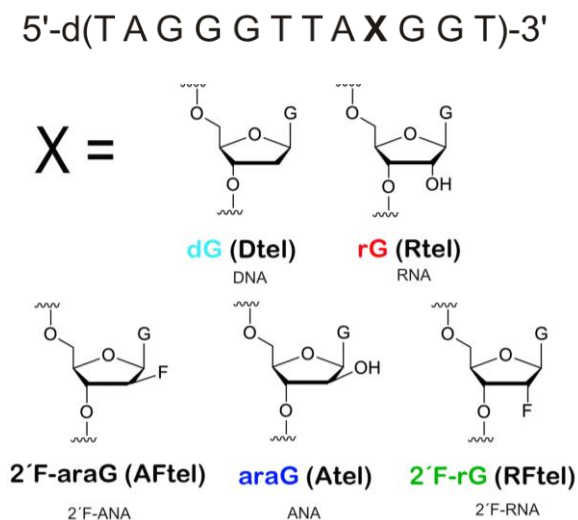


Figure 1: Oligonucleotide sequence and structures of the native (dG, rG) and chemically modified nucleotides (2'F-araG, araG and 2'F-rG).

9.2. RESULTS

9.2.1. G-quadruplex formation and thermodynamics

G-quadruplex formation and melting was monitored by Circular Dichroism (CD) and NMR spectroscopy in K^+ -containing solution (Figure 2, 3). The spectra of the unmodified sequence, **Dtel**, were characteristic of a mixture of parallel and antiparallel topologies (Figure 2). By contrast, **AFtel**, **Atel**, and **Rtel** fold mainly into a single G-quadruplex form, as indicated by the number of proton resonances; furthermore, the six sharp imino signals observed (10-12 ppm) and the data presented below are consistent with a symmetric, bimolecular G-quadruplex complex (Figure 2). By contrast, the imino region of **RFtel** shows very broad and weak signals, suggesting the presence of multiple species of very low stability (Figure 2).

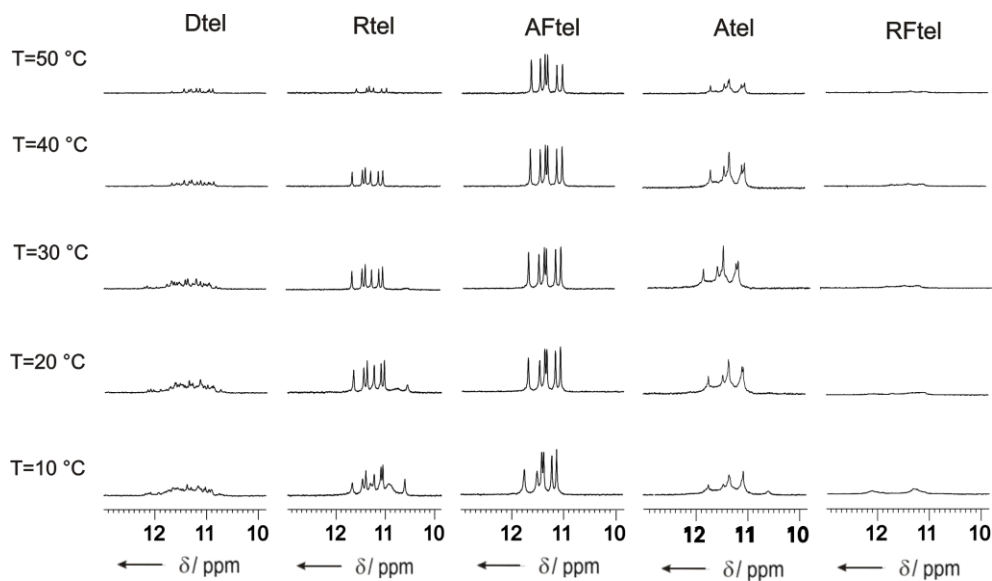


Figure 2: Imino region of 1H NMR proton spectra of the five modified telomeric sequences at different temperatures. Oligonucleotide concentration 0.7 mM (buffer conditions, 10 mM potassium phosphate, 5 mM potassium chloride, pH=7.0).

CD spectra of **AFtel**, **Atel**, and **Rtel** exhibit the distinctive features of a parallel quadruplex with a negative band at around 240 nm, and a positive one at 265 nm. In contrast, **Dtel** unmodified CD spectrum shows a positive pick at 265 nm and a

shoulder at 295 nm, characteristic of an equilibrium between parallel and antiparallel conformations in solution (Figure 3). The CD spectrum of **RFtel** is clearly different from the others, displaying low intensity bands, suggesting that it retains very little quadruplex structure, if at all, under the experimental conditions.

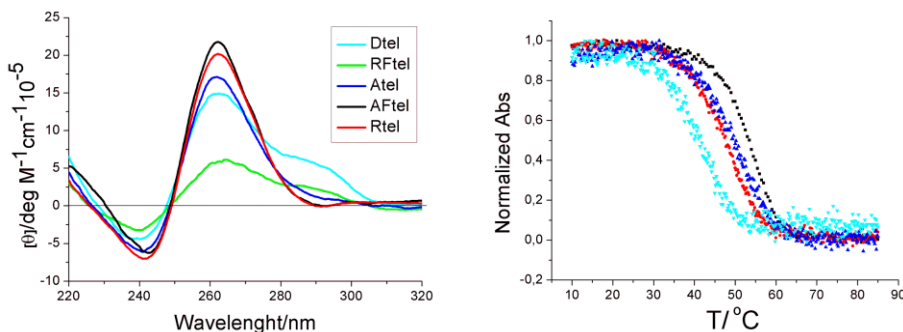


Figure 3: CD spectra and melting profiles of the five singly modified oligonucleotides.

Melting temperatures of all sequences were obtained from CD thermal denaturation experiments (Figure 3 and Table 1) and ^{19}F -NMR spectroscopy for **Aftel** (Figure 4). Surprisingly, in ^{19}F -NMR spectra, a very broad signal is observed at low temperatures. This large broadening effect was not observed in modified duplexes (chapter 5, 6 and 7), and may be provoked by the different quadruplex mobility compared to the single stranded form, or by the presence of a small population of higher order conformers in an intermediate equilibrium. A sharp signal is clearly observed at temperatures above 35 °C, most probably arising from the unfolded species.

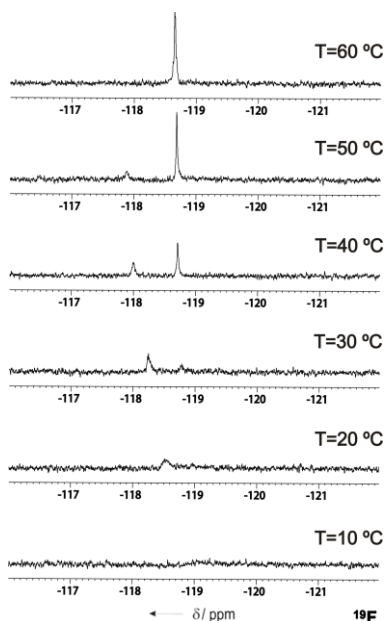


Figure 4: ^{19}F NMR spectra of **AFtel** at different temperatures.

Both arabinose modifications (**AFtel** and **Atel**) induce a strong stabilization, the effect being more pronounced for **AFtel** ($\Delta T_m = +9$ and $+12$ °C, respectively, relative to **Dtel**). In fact, **AFtel** exhibits the highest melting temperature of the entire series, exhibiting a T_m (53 °C) which rivals that of the all-RNA parallel quadruplex under similar conditions [r(UAGGGUUAGGGU); $T_m = 55$ °C]. The rG substitution is also stabilizing ($\Delta T_m = +6$ °C), in contrast to the 2'F-rG modification which is destabilizing (Figure 1). The overall trend observed was 2'F-araG>araG>rG>dG>>2'F-rG (Table 1).

Table 1: T_m values for DNA telomeric sequence (**Dtel**) and for singly-modified sequences^[a].

Name	Sequence (5'-3')	T_m (°C)	ΔT_m (°C)
Dtel	TAGGGTTAGGGT	41	0
Rtel	TAGGGTTA(rG)GGT	47	+6
Atel	TAGGGTTA(araG)GGT	50	+9
AFtel	TAGGGTTA(2'F-araG)GGT	53	+12
RFtel	TAGGGTTA(2'F-rG)GGT	-	-

^[a] Buffer conditions: 15 mM potassium phosphate, 5 mM KCl, pH=7.0; strand concentration 75uM.

9.2.2. NMR assignment and experimental constraints

Detailed structural information could be extracted from two-dimensional NMR spectra. In **AFtel**, **Atel** and **Rtel**, the number of spin systems in the NMR spectra indicated the formation of a symmetric structure. Sequential connectivities for the guanine residues within the central core could be traced via NOESY experiments (Figure 5 and Table A9.1 for **AFtel** chemical shifts). The almost identical chemical shifts in the three samples indicate that the structures are very similar and the differences are located in the neighbourhood of the modified step (Figure 6). H8-H1' NOE cross-peaks intensity indicates that all guanines are in *anti* glycosidic conformation (Figure 5).

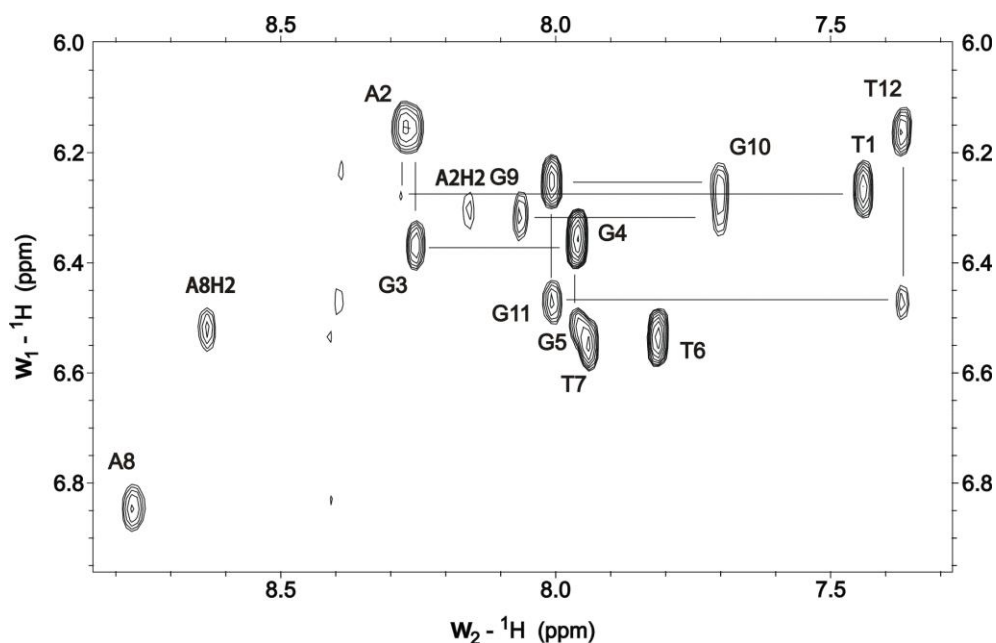


Figure 5: **AFtel** region of the NOESY spectrum in D_2O (250 ms mixing time). H1'-base assignment pathways are indicated and G9 label refers to 2'-F-araG modification.

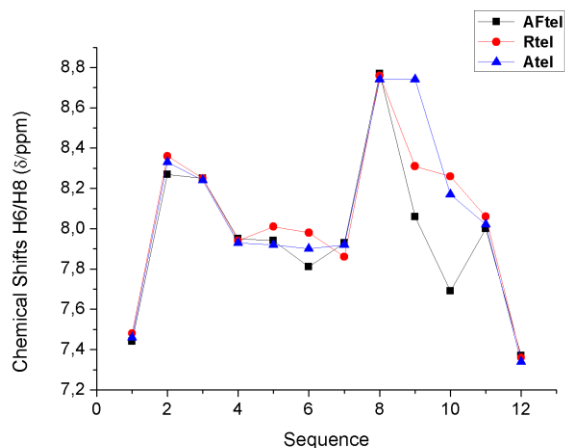


Figure 6: Aromatic H6/H8 chemical shifts vs sequence in **Atel**, **AFtel**, **Rtel**.

As expected for a symmetrical G-quadruplex with three tetrads, six imino protons were observed in the 10.5–12.0 ppm region of **AFtel**, **Rtel** and **Atel**. Although some peak overlapping was observed for **Atel**, two out of the six protons could be still observed after the samples were re-suspended in D₂O (Figure 7), suggesting that they are involved in the central tetrad

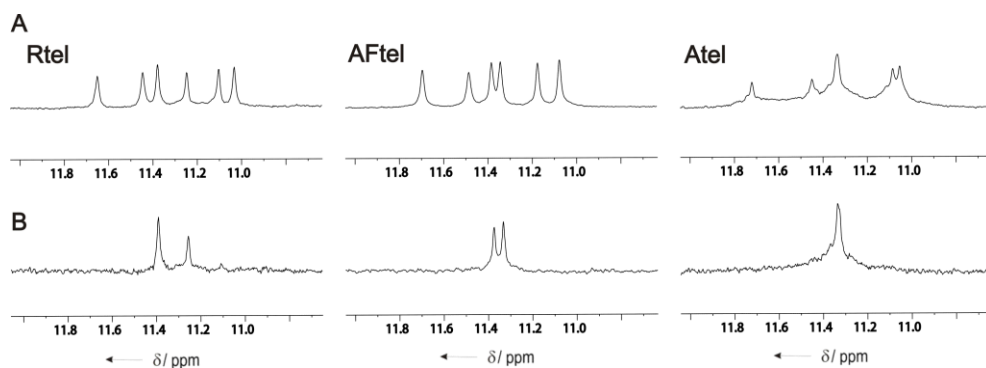


Figura 7: Parallel G-quadruplex NMR imino proton spectra of **Rtel**, **AFtel** and **Atel** modified telomeric sequence.

The exchangeable proton spectra of **AFtel** are very similar to that of the parallel conformer in the mixture formed by dTAGGGUTAGGGT; see Figure 3 in ref³¹. Consequently, exchangeable proton assignment was carried out on the basis of this structure³¹. These assignments were fully consistent with the NOESY spectra recorded in H₂O (Figure 8). The G-tetrad alignment could be identified from the NOESY spectra on the basis of the characteristic imino-H8 connectivity, being G3·G9·G3·G9, G4·G10·G4·G10 and G5·G11·G5·G11 (G9 = 2'F-araG; Figure 8) with the clockwise directionality shown in Figure 8b. Two TTA loops connect the tetrads, with the terminal residues (T1-A2 and T12) stacked on top and bottom of the G-quadruplex.

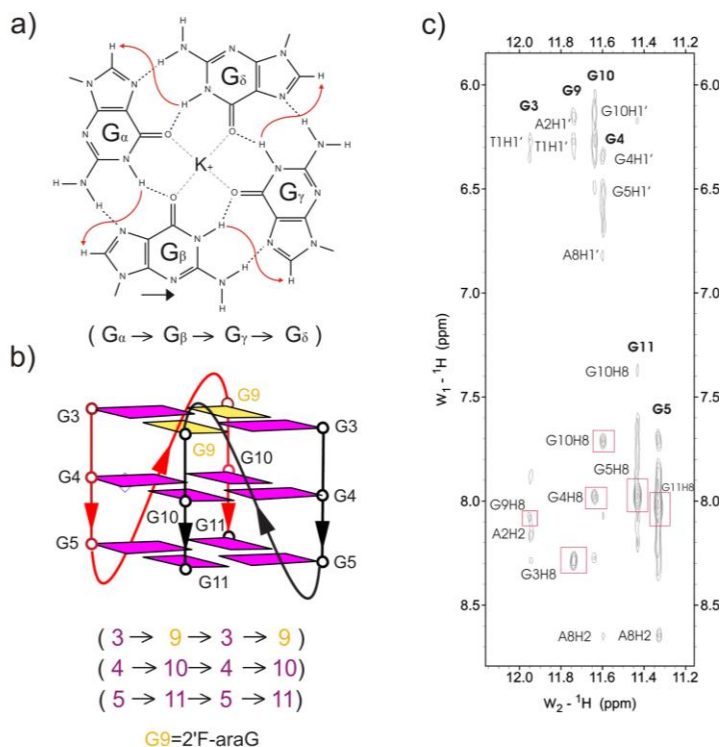


Figure 8: a) G-tetrads directionalities. b) Parallel propeller G-quadruplex scheme with all guanines in anti conformation. 2'F-araG nucleotides are shown in yellow and dG residues in magenta. c) **AFtel** imino region of the NOESY spectrum in H₂O (100 ms mixing time). The G-tetrad alignment was identified from the NOESY on the basis of the characteristic imino-H8 connectivity (indicated by red squares) around the G-tetrad.

The dG residues of **AFtel**, **Atel** and **Rtel** all adopt the South sugar conformation as determined from DQF-COSY derived J-coupling constants (Figure 9). The arabinose residues in **AFtel** and **Atel** adopt pure South sugar conformations (Figure 9), whereas rG (**Rtel**) favors the North conformation (data not shown), as expected. $J_{1'-2''}$, $J_{3'-2''}$ in **AFtel** and **Atel** are almost 0 Hz, characteristic of South conformation, while $J_{1'-2''}$ in rG9 is different than 0 Hz and, representing that the ribose sugar conformation is not entirely North.

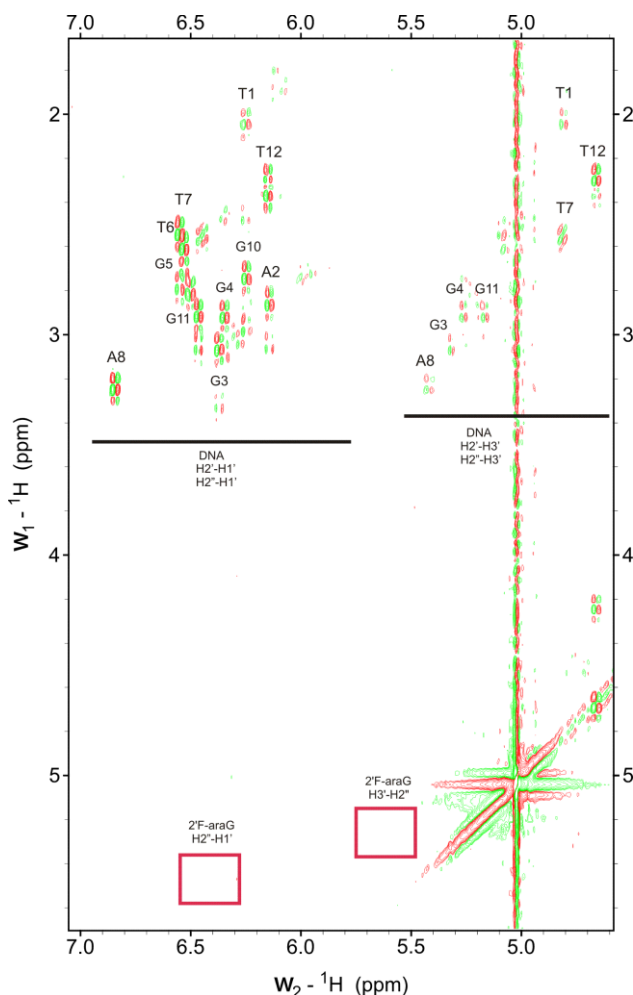


Figure 9: *AFtel* region of the DQF-COSY spectrum in D_2O , $H1'-H2''$, $H1'-H2'$, $H2'-H3'$, $H2''-H3'$ signals for DNA are at 2.5 ppm. $H1'-H2''$ and $H3'-H2''$ signals for $2'F\text{-araG}$ nucleotides are at 5.5 and 5.2 ppm indicated by red squares.

NMR-derived distances and torsion angle constraints together with restrained molecular dynamics were used to calculate the three-dimensional structure of **AFtel** (Table 2). Interproton distances for **AFtel** quadruplex were classified based on NOESY experiments run in H₂O (mixing times, 100 and 250 ms) and D₂O (mixing times, 150 and 250 ms). Qualitative limits of 1.8 Å and 5 Å were set in those distances where no quantitative analysis could be carried out, such as overlapping cross-peaks or those with a very weak intensity. No ¹⁹F-¹H distance constraints were included in the calculation, since ¹⁹F line broadening made acquisition of HOESY experiments impossible. In addition to this experimentally derived constraints, Hoogsteen hydrogen bond restraints and planarity restraints were imposed during structure calculations.

The G-quadruplex core is parallel, symmetric and very well defined (RMSD of 0.7 Å), with all guanines in the *anti* conformation (Figure 8 and 10). Continuous stacking of T1, A2 and G3 was supported by indicative NOEs, such as A2H2-G3H6 or T1CH3-G9H1. The same applies to T12 at the bottom of the G-core, which stacks over the G5-G11-G5-G11 tetrad. NMR data indicate that **Atel** and **Rtel** adopt similar structures (Figure 6). In conclusion, **AFtel**, **Atel** and **Rtel** adopt a parallel, propeller-like, bimolecular quadruplex structure.

Table 2: NMR restraints and structural statistics.

Experimental ¹ H- ¹ H distance constraints	
Total number	396
Intra-residue	288
Sequential	78
Long range	30
RMSD (Å)	
all heavy atoms	
- G-tetrad core	0.7 ± 0.2
- all residues	4.6 ± 0.6
all bases	
- G-tetrad core	0.3 ± 0.1
- all residues	2.2 ± 1.0
sum of violations (Å) (average and range)	7.7 (7.3-8.1)
max. violation (average and range)	0.23 (0.20-0.24)
NOE energy kcal/mol (average and range)	75.6 (73.3-77.3)

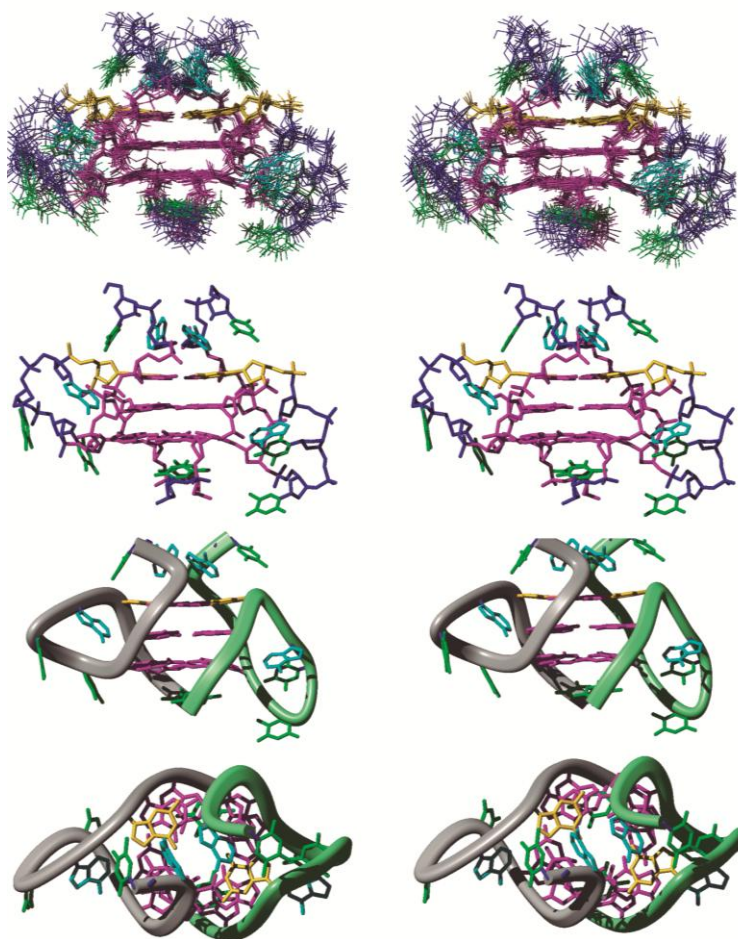


Figure 10: **AFtel** structure (PDB code: 2M1G) stereoviews of the 10 superimposed refined structures, the front view, and the front and top ribbon representation of the average structure.

9.3. DISCUSSION

The structure of **AFtel** was then compared with the crystallographic structure of $d^{Br}UAGGG^{Br}UTAGGGT$ (PDB ID: 1K8P) as a model of the DNA telomeric sequence¹⁴. The RMSDs between both structures are around 3.4 Å and 1.5 Å, for

all heavy atoms and the G-core, respectively. Superposition of the G-core residues in both structures indicates that the main differences reside around the 2'F-araG modification (Figure 11). Specifically, the 2'F-araG sugar residues (positions 9 in each subunit) are displaced towards the minor groove and consequently towards their 3'-neighbouring base (Figure 11, Table A9.2 and A9.3). This effect provokes a substantial decrease in the distance between the 2'-fluorine and H8 of the adjacent (3'-dG10) base (Figure 11). The 2'F..H8 distance for **AFtel** is around 2.8 Å, compared to 3.8 Å for the crystallographic structure (Figure 11). The close 2'F..H8 contact³² is indicative of a pseudohydrogen bond³³, as those found in 2'F-ANA:RNA hybrid duplexes (chapter 6)³⁴. For **AFtel**, the 2'F..H8-C angle is around 140° (Table A9.3), which is 20° higher than that observed in the unmodified dG-core. This value (140°) is similar to that observed in the 2'F-ANA:RNA hybrid (145°), and it is considered optimum for formation of F..H-X (X=N, O or C) pseudohydrogen bonds^{35,36}. The **AFtel** structure revealed an additional close contact between H2'' and O4' of the neighboring 3'-dG residue (2.9 Å in **AFtel** vs 4.4 Å in the crystal structure), with a C-H2''..O4' angle of 150° (versus 120° for 1K8P) (Figure 11 and Tables A9.4 and A9.5). Stabilizing FC-H2''..O interactions have previously been described in chapter 7 for fluorine modified duplexes and for nucleic acid³⁷ and protein-protein interactions^{38,39}. Therefore we hypothesize that the enhanced stability imparted by the 2'F-araG modification is due, at least in part, to 2'F..H8 and FC-H2''..O4' interactions, both of which are lacking in the native structure.

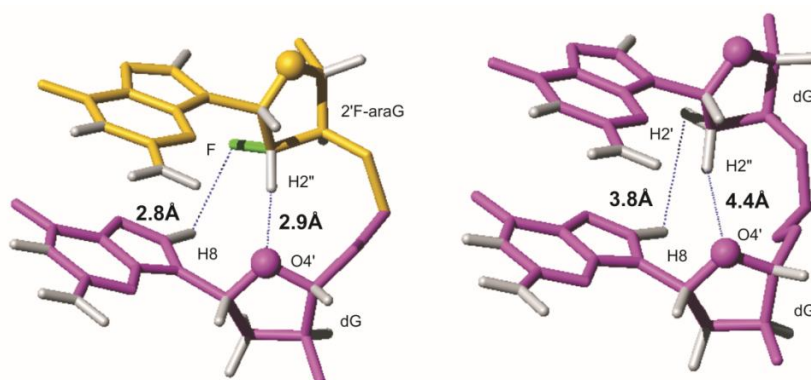


Figure 11: Distances between F_i-H8_{i+1} and $H2''_i-O4'_{i+1}$ in **AFtel** and $H2'_i-H8_{i+1}$ and $H2''_i-O4'_{i+1}$ in the crystallographic structure (1K8P). Guanines are shown in magenta for dG and yellow for 2'F-araG.

Based on the strong structural similarity, further computational calculations have shown that for **Atel**, the top-side 2'-OH of the araG nucleotide can easily be accommodated without incurring the steric penalty observed in duplex structures (chapter 5 and 6)^{28,34}. Furthermore, the arabinose 2'-OH is involved in favorable electrostatic interactions with the neighboring phosphates, as also seen for rG (**Rtel**) but not **RFtel** (Figure 12). This makes ara 2'-OH modifications particularly useful for applications where additional stabilization of a quadruplex and concomitant destabilization of a competing duplex structure is desired. The contrasting behavior of **Rtel** relative to **RFtel** is quite striking given the small structural differences between these nucleotides and the fact that 2'-F-RNA modifications are generally stabilizing⁴⁰⁻⁴². Whereas molecular dynamic simulations failed to detect spontaneous unfolding of the **RFtel** sequence, thermodynamic integration calculations, carried out by Guillem Portella and Modesto Orozco, indicated that the apparently innocuous change from 2'-OH to 2'-F generates a strong electrostatic repulsion between the fluorine group (Figure 12) and oxygen atoms in the backbone (around 26 kJ/mol per substitution).

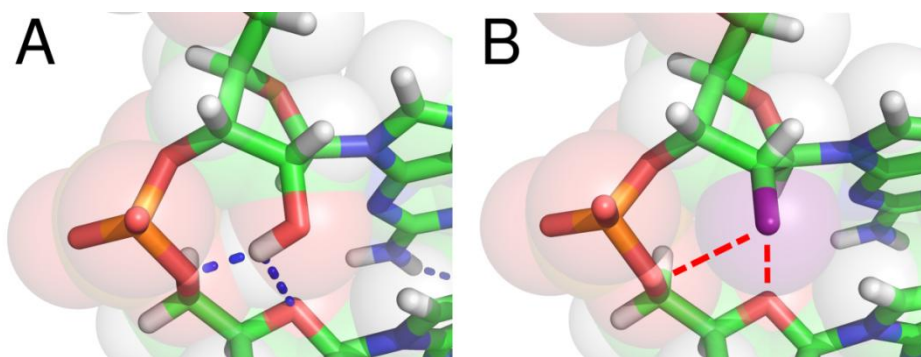


Figure 12: Snapshots of the most relevant configuration of the rG (A) and 2'-F-rG (B) in the G-quadruplex as found in the molecular dynamic simulations.

9.4. CONCLUSION

In conclusion, arabinose and ribose substitutions, and in particular the 2'F-ara substitution, can be used to stabilize the biologically relevant propeller parallel G-quadruplex form over competing conformers, allowing NMR structural determination of this particularly significant nucleic acid structure. 2'F-araG allowed the greatest stabilization (+12 °C) of the parallel quadruplex, as a result of electrostatic (FC-H2''..O4') and pseudohydrogen bonding (2'F..H8) stabilizing interactions. Sugar/phosphate interactions stabilize the ara and ribo substitutions, whereas 2'F-rG substitution provoked a dramatic destabilization (disruption) of the quadruplex structure. Thus a single conservative oligonucleotide chemical modification can exert remarkable control over G-quadruplex formation, enabling significant structural insight into the nature of the propeller parallel G-quadruplex.

BIBLIOGRAPHY

- (1) Sandell, L. L.; Zakian, V. A. *Loss of a yeast telomere: Arrest, recovery, and chromosome loss* **Cell** (1993), 75, 729.
- (2) De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J.-F.; Mergny, J.-L. *Targeting telomeres and telomerase* **Biochimie** (2008), 90, 131.
- (3) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. *Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo* **Nat Struct Mol Biol** (2005), 12, 847.
- (4) Nikan, M.; Sherman, J. C. *Template-Assembled Synthetic G-Quartets (TASQs)* **Angew. Chem. Int. Ed. Engl.** (2008), 47, 4900.
- (5) So, H.-M.; Won, K.; Kim, Y. H.; Kim, B.-K.; Ryu, B. H.; Na, P. S.; Kim, H.; Lee, J.-O. *Single-Walled Carbon Nanotube Biosensors Using Aptamers as Molecular Recognition Elements* **J. Am. Chem. Soc.** (2005), 127, 11906.
- (6) Hsu, C.-L.; Chang, H.-T.; Chen, C.-T.; Wei, S.-C.; Shiang, Y.-C.; Huang, C.-C. *Highly Efficient Control of Thrombin Activity by Multivalent Nanoparticles* **Chem. Eur. J.** (2011), 17, 10994.
- (7) Ge, B.; Huang, Y. C.; Sen, D.; Yu, H.-Z. *A Robust Electronic Switch Made of Immobilized Duplex/Quadruplex DNA* **Angew. Chem. Int. Ed. Engl.** (2010), 49, 9965.
- (8) Huang, Y. C.; Cheng, A. K. H.; Yu, H.-Z.; Sen, D. *Charge Conduction Properties of a Parallel-Stranded DNA G-Quadruplex: Implications for Chromosomal Oxidative Damage* **Biochemistry** (2009), 48, 6794.
- (9) Huang, Y. C.; Sen, D. *A Contractile Electronic Switch Made of DNA* **J. Am. Chem. Soc.** (2010), 132, 2663.
- (10) Yang, D.; Okamoto, K. *Structural insights into G-quadruplexes: towards new anticancer drugs* **Future Med. Chem.** (2010), 2, 619.
- (11) Phan, A. T.; Kuryavyi, V.; Ma, J.-B.; Faure, A.; Andréola, M.-L.; Patel, D. J. *An interlocked dimeric parallel-stranded DNA quadruplex: A potent inhibitor of HIV-1 integrase* **Proc. Natl. Acad. Sci. U.S.A.** (2005), 102, 634.
- (12) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Quadruplex DNA: sequence, topology and structure* **Nucleic Acids Res.** (2006), 34, 5402.
- (13) Phan, A. T. *Human telomeric G-quadruplex: structures of DNA and RNA sequences* **FEBS J.** (2010), 277, 1107.
- (14) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Crystal structure of parallel quadruplexes from human telomeric DNA* **Nature** (2002), 417, 876.
- (15) Phan, A. T.; Luu, K. N.; Patel, D. J. *Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K⁺ solution* **Nucleic Acids Res** (2006), 34, 5715.

- (16) Azzalin, C. M.; Reichenbach, P.; Khoraiuli, L.; Giulotto, E.; Lingner, J. *Telomeric Repeat-Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends* **Science** (2007), 318, 798.
- (17) Stefan, S.; Maria, A. B. *Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II* **Nat. Cell. Biol.** (2007), 10, 228.
- (18) Martadinata, H.; Phan, A. T. *Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K⁺ solution* **J. Am. Chem. Soc.** (2009), 131, 2570.
- (19) Collie, G. W.; Haider, S. M.; Neidle, S.; Parkinson, G. N. *A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex* **Nucleic Acids Res** (2010), 38, 5569.
- (20) Tang, C. F.; Shafer, R. H. *Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecular recognition in guanine quadruplexes* **J. Am. Chem. Soc.** (2006), 128, 5966.
- (21) Karsisiotis, A.; Webba da Silva, M. *Structural Probes in Quadruplex Nucleic Acid Structure Determination by NMR* **Molecules** (2012), 17, 13073.
- (22) Peng, C. G.; Damha, M. J. *G-quadruplex induced stabilization by 2'-deoxy-2'-fluoro-D-arabinonucleic acids (2'-F-ANA)* **Nucleic Acids Res** (2007), 35, 4977.
- (23) Lech, C. J.; Li, Z.; Heddi, B.; Phan, A. T. *2'-F-ANA-guanosine and 2'-F-guanosine as powerful tools for structural manipulation of G-quadruplexes* **Chem. Commun.** (2012), 48, 11425.
- (24) Webba da Silva, M.; Trajkovski, M.; Sannohe, Y.; Ma'ani Hessari, N.; Sugiyama, H.; Plavec, J. *Design of a G-quadruplex topology through glycosidic bond angles* **Angew. Chem. Int. Ed. Engl.** (2009), 48, 9167.
- (25) Dominick, P. K.; Jarstfer, M. B. *A Conformationally Constrained Nucleotide Analogue Controls the Folding Topology of a DNA G-Quadruplex* **J. Am. Chem. Soc.** (2004), 126, 5050.
- (26) Doluca, O.; Withers, J. M.; Filichev, V. V. *Molecular Engineering of Guanine-Rich Sequences: Z-DNA, DNA Triplexes, and G-Quadruplexes* **Chem. Rev.** (2013), DOI:10.1021/cr300225q.
- (27) Sapse, A. M.; Snyder, G. *Ab initio studies of the antiviral drug 1-(2-fluoro-2-deoxy-beta-D-arabinofuranosyl) thymine* **Cancer Invest.** (1985), 3, 115.
- (28) Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; Gonzalez, C. *The solution structure of double helical arabino nucleic acids (ANA and 2'-F-ANA): effect of arabinoses in duplex-hairpin interconversion* **Nucleic Acids Res** (2012), 40, 9329.
- (29) Berger, I.; Tereshko, V.; Ikeda, H.; Marquez, V. E.; Egli, M. *Crystal structures of B-DNA with incorporated 2'-deoxy-2'-fluoro-arabino-furanosyl thymine: Implications of conformational preorganization for duplex stability* **Nucleic Acids Res.** (1998), 26, 2473.
- (30) Watts, J. K.; Damha, M. J. *2'-F-Arabinonucleic acids (2'-F-ANA) — History, properties, and new frontiers* **Can. J. Chem.** (2008), 86, 641.
- (31) Phan, A. T.; Patel, D. J. *Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics* **J. Am. Chem. Soc.** (2003), 125, 15021.
- (32) Li, F.; Sarkhel, S.; Wilds, C. J.; Wawrzak, Z.; Prakash, T. P.; Manoharan, M.; Egli, M. *2'-Fluoroarabino- and Arabinonucleic Acid Show Different Conformations, Resulting in Deviating RNA Affinities and Processing of Their Heteroduplexes with RNA by RNase H* **Biochemistry** (2006), 45, 4141.
- (33) Anzahaee, M. Y.; Watts, J. K.; Alla, N. R.; Nicholson, A. W.; Damha, M. J. *Energetically Important C-H...F-C Pseudohydrogen Bonding in Water: Evidence and Application to Rational Design of Oligonucleotides with High Binding Affinity* **J. Am. Chem. Soc.** (2010), 133, 728.
- (34) Watts, J. K.; Martín-Pintado, N.; Gómez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; González, C.; Damha, M. J. *Differential stability of 2'-F-ANA•RNA and ANA•RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility* **Nucleic Acids Res.** (2010), 38, 2498.
- (35) Mehta, G.; Sen, S. *Probing Fluorine Interactions in a Polyhydroxylated Environment: Conservation of a C-F...H-C Recognition Motif in Presence of O-H...O Hydrogen Bonds* **Eur. J. Org. Chem.** (2010), 2010, 3387.
- (36) Dalvit, C.; Vulpatti, A. *Intermolecular and intramolecular hydrogen bonds involving fluorine atoms: implications for recognition, selectivity, and chemical properties* **ChemMedChem** (2012), 7, 262.
- (37) Berger, I.; Egli, M.; Rich, A. *Inter-strand C-H...O hydrogen bonds stabilizing four-stranded intercalated molecules: stereoelectronic effects of O4' in cytosine-rich DNA* **Proc. Natl. Acad. Sci. U.S.A.** (1996), 93, 12116.
- (38) Jiang, L.; Lai, L. *CH...O Hydrogen Bonds at Protein-Protein Interfaces* **J. Biol. Chem.** (2002), 277, 37732.
- (39) Horowitz, S.; Yesselman, J. D.; Al-Hashimi, H. M.; Trievel, R. C. *Direct Evidence for Methyl Group Coordination by Carbon-Oxygen Hydrogen Bonds in the Lysine Methyltransferase SET7/9* **J. Biol. Chem.** (2011), 286, 18658.
- (40) Pallan, P. S.; Greene, E. M.; Jicman, P. A.; Pandey, R. K.; Manoharan, M.; Rozners, E.; Egli, M. *Unexpected origins of the enhanced pairing affinity of 2'-fluoro-modified RNA* **Nucleic Acids Res** (2011), 39, 3482.
- (41) Deleavey, G. F.; Watts, J. K.; Alain, T.; Robert, F.; Kalota, A.; Aishwarya, V.; Pelletier, J.; Gewirtz, A. M.; Sonenberg, N.; Damha, M. J. *Synergistic effects between analogs of DNA and RNA improve the potency of siRNA-mediated gene silencing* **Nucleic Acids Res.** (2010), 38, 4547.

- (42) Patra, A.; Paolillo, M.; Charisse, K.; Manoharan, M.; Rozners, E.; Egli, M. *2'-Fluoro RNA Shows Increased Watson–Crick H-Bonding Strength and Stacking Relative to RNA: Evidence from NMR and Thermodynamic Data* **Angew. Chem. Int. Ed. Engl.** (2012), *51*, 11863.

Discusión general

A lo largo de esta tesis hemos realizado un estudio detallado mediante RMN de diversos oligonucleótidos modificados en la posición C2' del azúcar. Concretamente nos hemos centrado en el estudio de nucleótidos fluorados en la posición C2' de la furanosa (2'F-ANA, 2'F-RNA y dFdC) y sus correspondientes análogos hidroxilados (ANA y RNA). El principal interés de estas modificaciones es que, aunque aparentemente se trata de variaciones pequeñas en la composición química del DNA y RNA, son capaces de producir efectos drásticos en la conformación y estabilidad de los ácidos nucleicos. La determinación de las estructuras de estas moléculas a nivel atómico nos ha permitido caracterizar diversos tipos de interacciones en las que están involucrados los grupos 2'OH o 2'F. Estas interacciones pueden ser favorable o desfavorables, dependiendo del contexto estructural en el que se encuentran. En esta tesis, se han estudiado estas modificaciones en diversos tipos de dúplex (tipo A, B o intermedios), así como en horquillas y en cuádruplex de guaninas.

Los nucleótidos de 2'F-RNA tienden a adoptar una conformación tipo norte (C3'-*endo*) y a formar estructuras de dúplex tipo A, muy similares a las de su análogo hidroxilado, el RNA, pero mucho más estables. Diversos estudios han atribuido dicha estabilización a un efecto de polarización del flúor sobre la nucleobase, que provoca una interacción de apilamiento más favorable entre nucleótidos 2'F-RNA contiguos y la formación de pares de bases entre 2'F-RNA complementarios más estables que los formados en el RNA. La determinación estructural de dúplex quiméricos 2'F-RNA/2'F-ANA nos ha llevado a formular una hipótesis alternativa. La comparación de las estructuras con secuencias alternas y continuas de nucleótidos 2'F-RNA y 2'F-ANA nos ha permitido detectar la presencia de interacciones electrostáticas favorables entre el protón H2'/H2'' con átomos de oxígeno del esqueleto azúcar-fosfato de los nucleótidos circundantes. La electronegatividad del flúor cambia la distribución de cargas en la ribosa y polariza el enlace 2'C-H, favoreciendo la formación de enlaces de hidrógeno no convencionales (FC-H2'_i..O4'_{i+1}) mucho más intensos que los encontrados en RNA (HOC-H2'_i..O4'_{i+1}). La ruptura de este tipo de interacciones en pasos 2'F-RNA/2'F-ANA de regiones alternas 2'F-ANA/2'F-RNA/2'F-ANA, es la responsable

de la desestabilización de dúplex con secuencias alternas frente a aquellos con tramos continuos de 2'F-RNA o 2'F-ANA.

Los nucleótidos de 2'F-ANA, por el contrario, tienden a adoptar una conformación sur (C2'-*endo*). En esta tesis se ha llevado a cabo la primera determinación estructural de un dúplex con nucleótidos 2'F-ANA en las dos cadenas y se ha concluido que forman estructuras tipo B. En este tipo de estructuras, la conformación sur del azúcar permite la formación de interacciones estabilizantes 2'F..H8 intranucleotídicas. Por otra parte, en dúplex híbridos 2'F-ANA:RNA y 2'F-ANA:2'F-RNA, hemos observado que los nucleótidos de 2'F-ANA son capaces de adaptarse a una conformación sudeste, que permite la formación de varias interacciones estabilizantes. Así, en dúplex 2'F-ANA:RNA se forman interacciones 2'F..H8 secuenciales en pasos pirimidina-purina. La geometría en estos pasos permite el acercamiento del átomo de flúor al H8 de la purina secuencial formándose pseudoenlaces de hidrógeno. Sorprendentemente, este tipo de interacciones no se favorecen en el dúplex híbrido 2'F-ANA:2'F-RNA. La usencia de las interacciones 2'F..H8 es debida a que los nucleótidos de 2'F-RNA fuerzan una estructura tipo A, mientras que en los híbridos 2'F-ANA:RNA la estructura es intermedia entre A y B. En los primeros, el ángulo glicosídico de las bases es 130-140° impidiendo una geometría óptima para la formación de la interacción 2'F..H8. Hay que destacar que la conformación sudeste de los azúcares favorece, en ambas estructuras, las interacciones $FC-H2'_{i..}O4'_{i+1}$ entre nucleótidos de 2'F-ANA, aunque la geometría no parece ser tan favorable como en el caso de 2'F-RNA. El resultado de todas estas interacciones es una gran estabilización de este tipo de dúplex con respecto a los de DNA y RNA.

El caso de los nucleótidos con dos átomos de flúor en la posición C2' (dFdC) es interesante, ya que estudios previos muestran que esta modificación es desestabilizante cuando se inserta, tanto en dúplex tipo B (DNA:DNA) como en tipo A (RNA:RNA). Por ello, decidimos determinar el efecto de esta modificación en dúplex híbridos DNA:RNA de tipo intermedio. Sorprendentemente, el dFdC en este tipo de dúplex prácticamente no produce cambios en su estabilidad, aunque las estructuras obtenidas por RMN claramente muestran que esta modificación

produce distorsiones en los pares de bases. La desestabilización provocada por estas distorsiones parece estar compensada por la formación de las interacciones 2'F..H8 secuenciales en el paso pirimidina-purina

De particular interés, es la comparación de estructuras análogas con 2'F (2'F-ANA) o con 2'-OH (ANA). En esta tesis hemos observado que, en conformaciones de doble hélice, tanto ANA como 2'F-ANA forman estructuras muy similares pero tienen un efecto opuesto en la estabilidad. La determinación estructural del dúplex ANA:ANA y ANA:RNA nos ha permitido concluir que, en ambos casos, el efecto desestabilizante se debe a contactos desfavorables entre el grupo hidroxilo y el H8 de las bases. Esta interacción desestabilizante es intranucleotídica en dúplex ANA:ANA (tipo B) e internucleotídica en dúplex híbridos ANA:RNA (tipo intermedio).

Cuando la estructura permite evitar los contactos desfavorables con el grupo 2'-OH, ANA deja de ser una sustitución desestabilizante. Por ejemplo, en estructuras de cadena sencilla, ANA permite la formación de enlaces de hidrógeno entre el grupo hidroxilo y los oxígenos del esqueleto, favoreciendo así las estructuras tipo horquilla con los nucleótidos de ANA desapareados. Lo mismo ocurre en las estructuras de cuádruplex de guaninas paralelos estudiadas en esta tesis, donde la nucleobase de araG se dispone en una conformación que no provoca contactos desfavorables con el grupo 2'-OH. Éste, a su vez, forma enlaces de hidrógeno con los fosfatos del esqueleto, lo que provoca que ANA sea una sustitución muy estabilizante en este tipo de estructuras.

El hecho de que diferentes modificaciones provoquen efectos diferentes e incluso opuestos dependiendo de la estructura secundaria no está limitado al caso de ANA. Nuestros estudios con la secuencia del telómero humano muestran que los nucleótidos de 2'F-ANA, ANA y RNA estabilizan el cuádruplex paralelo, mientras los nucleótidos de 2'F-RNA desestabilizan la formación del cuádruplex. La estabilización más drástica viene dada por 2'F-ANA. La determinación de la estructura tridimensional del cuádruplex que contiene esta sustitución nos ha permitido concluir que la estabilización se debe a que, en la estructura de cuádruplex paralelo, tienen lugar simultáneamente tanto las interacciones 2'F..H8

secuenciales como las $FC-H2''_i..O4'_{i+1}$. La inserción de nucleótidos de RNA, en el cuádruplex, estabilizan la estructura secundaria gracias a la interacción de los grupos hidroxilo con el esqueleto del nucleótido secuencial, de manera análoga a como ocurre con ANA. Por el contrario, los nucleótidos de 2'F-RNA provocan una fuerte desestabilización debido a interacciones electrostáticas desfavorables entre el 2'F y los oxígenos del esqueleto del nucleótido secuencial.

Estos resultados ponen de manifiesto que el efecto de las modificaciones en C2' en la estabilidad de los ácidos nucleicos es dependiente de la estructura secundaria que éstos adoptan. Éste es probablemente uno de los resultados más importantes de esta tesis y añade un mayor interés a este tipo de sustituciones. Estas modificaciones no sólo son útiles para construir oligonucleótidos resistentes a nucleasas para terapias antisentido o RNAi, sino que pueden ser una herramienta muy útil para modular la estructura secundaria de ácidos nucleicos. Por ejemplo, la incorporación de nucleótidos de ANA en un oligonucleótido rico en guaninas permite la desestabilización de estructuras tipo dúplex (y presumiblemente de tríplex) frente a la estructura de cuádruplex. Por el contrario, la incorporación de 2'F-RNA en el mismo oligonucleótido provocará la desestabilización del cuádruplex frente a la estructura de dúplex o tríplex, propiedad muy deseable en numerosas aplicaciones.

Conclusiones

- Se ha estudiado el comportamiento del oligonucleótido autocomplementario **gap(FA)**, que combina nucleótidos de ANA y 2'F-ANA en la secuencia. Se ha observado que, en disolución, coexisten dos especies en equilibrio. Una de estas especies es un dúplex, con todos los nucleótidos formando pares de bases. La segunda especie es una horquilla, donde sólo los nucleótidos 2'F-ANA están formando pares de bases y los ANA están desapareados.
- Se ha resuelto la estructura tridimensional del dúplex **gap(FA)**. El análisis detallado de este dúplex muestra una conformación global tipo B con todos los azúcares en sur (C2'-*endo*). En los nucleótidos de 2'F-ANA, esta conformación del anillo favorece las interacciones 2'F..H8 intranucleotídicas, que contribuyen a la estabilización del dúplex. Por otra parte, el grupo hidroxilo de los nucleótidos de ANA provoca interacciones desfavorables con su propia base, lo que hace que ésta se desplace impidiendo una geometría óptima en los pares ANA:ANA. Estos dos efectos explican la gran diferencia en la estabilidad de los dúplex ANA:ANA y 2'F-ANA:2'F-ANA.
- Se ha modelizado la estructura de la horquilla ("*stem-loop*") formada por la secuencia **gap(FA)**. El "*stem*" de nucleótidos de 2'F-ANA mantiene una estructura tipo B muy similar a la observada en el dúplex. Los nucleótidos de ANA, por el contrario, se encuentran en el "*loop*" y mantienen una conformación sur (C2'-*endo*). Esta conformación del azúcar permite, tanto la mejor hidratación de los grupos hidroxilo, como la formación de enlaces de hidrógeno entre éstos y los oxígenos del esqueleto. Estas interacciones hacen que los arabino nucleótidos contribuyan a estabilizar estructuras tipo horquilla.
- Se ha determinado la estructura tridimensional de tres dúplex híbridos DNA:RNA (**DR**), 2'F-ANA:RNA (**FR**) y ANA:RNA (**AR**), sustratos de la *RNase H*. Los tres muestran una estructura global intermedia entre las formas A y B canónicas.
- El análisis detallado de estas estructuras muestra sutiles diferencias tanto a nivel de los azúcares como de las interacciones que en éstas se producen. Mientras que en el DNA los azúcares se encuentran en equilibrio entre norte y sur (C3'-

endo/C2'-*endo*), los 2'-F-ANA están en una conformación sudeste (O4'-*endo*), y los ANA en sur (C2'-*endo*).

- La estructura tridimensional de los híbridos 2'-F-ANA:RNA y ANA:RNA muestra que las interacciones que se producen en estas dos estructuras son diferentes y con efectos opuestos. Mientras en la cadena de 2'-F-ANA se forman interacciones favorables, como son los pseudoenlaces de hidrógeno 2'-F..H8 y las interacciones FC-H2''_i..O4'/O5'_{i+1}, en el caso de ANA, los contactos de van der Waals desfavorables entre el grupo hidroxilo y el H8 del nucleótido secuencial desestabilizan el dúplex. Estos efectos son los responsables de la mayor estabilidad de los dúplex híbridos 2'-F-ANA:RNA con respecto a los ANA:RNA.

- Se ha determinado la estructura tridimensional de dos dúplex quiméricos 2'-F-RNA/2'-F-ANA con secuencias contiguas y alternas de ambos nucleótidos (**H** y **S**, respectivamente). Se ha observado una estructura global intermedia en ambos casos, en las que los azúcares de 2'-F-ANA se mantienen en conformaciones sudeste (O4'-*endo*/C2'-*endo*), mientras que los nucleótidos de 2'-F-RNA están en norte (C3'-*endo*) en secuencias continuas, o noreste si se encuentran entre nucleótidos de 2'-F-ANA (**S**).

- Se ha demostrado que el átomo de flúor en los nucleótidos 2'-F-ANA y 2'-F-RNA polariza el enlace C-H2'/H2'' y favorece las interacciones FC-H2'/H2''_i..O_{i+1} con el esqueleto del nucleótido 3'-secuencial. Estas interacciones tienen un efecto muy significativo en la estabilización de las estructuras secundarias que contienen este tipo de modificaciones.

- Las interacciones FC-H2'/H2''_i..O_{i+1} permiten explicar la diferente estabilidad de dúplex con secuencias contiguas y alternas de nucleótidos 2'-F-RNA y 2'-F-ANA. Se ha demostrado que la menor estabilidad las secuencias alternas se debe a que, en secuencias 2'-F-ANA/2'-F-RNA/2'-F-ANA, la conformación de 2'-F-RNA no permite la formación de la interacción C-H_i..O4'_{i+1} con el nucleótido 2'-F-ANA siguiente.

- Se ha determinado la estructura tridimensional de un dúplex híbrido DNA:RNA con un nucleótido dFdC insertado en la cadena de DNA. Se ha observado que el

nucleótido modificado adopta una conformación sudeste que provoca distorsiones en los pares de bases adyacentes a dicha modificación. Por otro lado, esta conformación del azúcar favorece la formación del pseudoenlace secuencial 2'F..H8 que compensa parcialmente la desestabilización provocada por las distorsiones de la estructura.

- Se ha estudiado el efecto de nucleótidos modificados en C2', en la estabilidad y conformación de la secuencia telomérica de DNA (**Dtel**). En disolución, este oligonucleótido adopta estructuras de cuádruplex paralelo y antiparalelo que coexisten en equilibrio. La inserción de un 2'F-ANA, ANA o RNA desplaza el equilibrio hacia el cuádruplex paralelo, mientras un nucleótido de 2'F-RNA es capaz de desestabilizar drásticamente la formación del cuádruplex.

- Se ha determinado que la modificación de 2'F-ANA es la que más estabiliza el cuádruplex paralelo, seguida de la estabilización con ANA y finalmente con RNA. Este resultado pone de manifiesto que el efecto de este tipo de modificaciones en la estabilidad de los ácidos nucleicos es dependiente de la estructura secundaria que éstos adoptan.

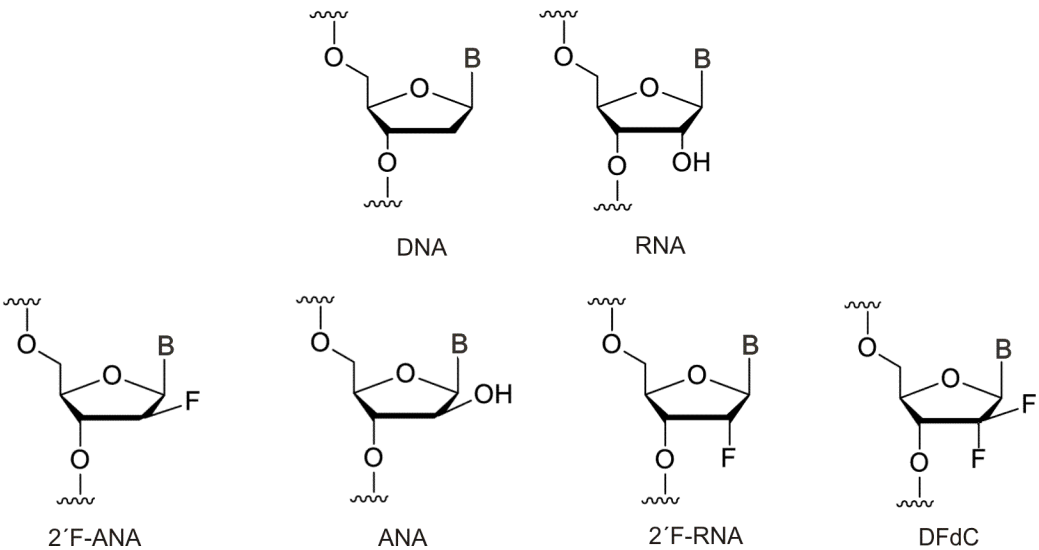
- Se ha determinado la estructura tridimensional del cuádruplex paralelo formado por la secuencia telomérica **AFtel**, con un nucleótido de 2'F-ANA. Estructuralmente se ha demostrado que esta estabilización está provocada, en parte, por las interacciones favorables 2'F..H8 y C-H2''..O del nucleótido de 2'F-ANA con el nucleótido 3'-secuencial.

- Se ha demostrado que la estabilización del cuádruplex al insertar nucleótidos de ANA y RNA se debe, principalmente, a interacciones favorables del grupo hidroxilo con el esqueleto. Por otro lado, la desestabilización provocada por el 2'F-RNA se debe a interacciones desestabilizantes del flúor con los oxígenos del esqueleto.

Estructuras determinadas experimentalmente a lo largo de esta tesis

Secuencia ^a	Nombre	PDB ID	Capítulo
5'-CGCGAAUUCGCG-3' 5'-GCGCUUAAGCGC-3'	Gap(FA)	2LSC	Capítulo 5
5'-GCTATAATGG-3' 5'-CGAUUUUACC-3'	FR	2KP4	Capítulo 6
5'-GCUAUAAUGG-3' 5'-CGAUUAUACC-3'	AR	2KP3	Capítulo 6
5'-CGCGAATTCGCG-3' 5'-GCGCTTAAGCGC-3'	H	2M84	Capítulo 7
5'-CGCGAATUCGCG-3' 5'-GCGCUTAAGCGC-3'	S		Capítulo 7
5'-CTACACGG-3' 5'-GAUGUGCC-3'	[CON]	Por enviar	Capítulo 8
5'-CTACACGG-3' 5'-CAUGUGCC-3'	[GEM]	Por enviar	Capítulo 8
Dimero 5'-TAGGGTTAGGGT-3'	AFtel	2M1G	Capítulo 9

^aLeyenda: DNA, RNA, 2'F-ANA, ANA, 2'F-RNA y dFdC.



Annex I

Material and methods

Oligonucleotide synthesis and purification

Oligonucleotides were synthesized from phosphoramidite precursors using standard solid-phase methods¹. Purification of oligonucleotides was carried out using reverse phase HPLC and all masses were verified by ESI-MS or MALDI-TOF. Small triethylammonium signals were visible in the NMR spectra from the ion-pairing reagent, but did not interfere with an important region of the NMR spectrum.

Oligonucleotides used in chapter 5, 6, 7 and 9 were synthesized in Prof. Masad Damha's laboratory from McGill University in Montreal. Oligonucleotides used in chapter 8 were synthesized in Prof. Ramón Eritja's laboratory from IQAC-CSIC in Barcelona.

UV and CD melting experiments and derivation of thermodynamic parameters

UV and CD data were obtained on a Varian Cary 5000 UV-VIS spectrophotometer and a JACSO J-810 circular dichroism spectrometer respectively. Both equipped with a Peltier temperature controller. Duplex concentration was 2 μM (4 μM total concentration of strands) and G-quadruplex oligonucleotide concentration was 75 μM . After heating to 90°, samples were cooled to room temperature at a rate of 0.3 (°C/min) and then refrigerated overnight. Absorbance values were recorded after equilibration as the temperature was increased in 0.5 °C steps at 1-min intervals in duplexes and 0.3 °C steps at 1-min intervals in G-quadruplexes. T_m values were calculated using the baseline method since assignment of baselines was clear in most cases. There is no hysteresis in duplexes, while in G-quadruplex, despite the slow scan rate, some hysteresis was observed in all cases suggesting slow kinetics of formation/dissociation. This is a common effect in multimeric G-quadruplexes, which makes that thermodynamics parameters should be consider with caution.

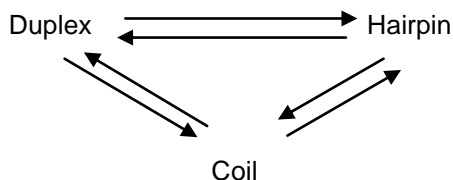
NMR experiments

High concentration samples (0.6-0.8 mM) were suspended in 300 or 500 μL of either D_2O or $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1. Duplexes buffer was 25 mM sodium phosphate, 100 mM NaCl, pH 7, and G-quadruplexes buffer was 10 mM potassium phosphate, 5

mM KCl, pH 7. NMR spectra were acquired in Bruker spectrometers operating at 600, 700 or 800 MHz, and processed with Topspin software. ^{19}F and ^1H 1D melting experiments, DQF-COSY, TOCSY and NOESY experiments were recorded in D_2O and $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1. The NOESY spectra were acquired with mixing times of 50, 100, 150, 250 and 300 ms, and the TOCSY spectra were recorded with standard MLEV-17 spin-lock sequence, and 80-ms mixing time. For highly concentrated samples NOESY spectra in H_2O were acquired with 50, 100, 150 and 250 ms mixing times. For 2D experiments in H_2O , water suppression was achieved by including a WATERGATE² module in the pulse sequence prior to acquisition. Two-dimensional experiments in D_2O were carried out at temperatures ranging from 5 °C to 35 °C, whereas spectra in H_2O were recorded at 5-10 °C to reduce the exchange with water. ^{19}F resonances were assigned, in most cases, from ^{19}F detected HOESY spectra³ ($\tau_m = 200$ ms) and ^1H - ^{19}F HETCOR. The spectral analysis program Sparky⁴ was used for semiautomatic assignment of the NOESY cross-peaks and quantitative evaluation of the NOE intensities.

NMR melting experiments in chapter 5

The dissociation of both structures in **gap(FA)** i.e., the double stranded and the hairpin structure is described with the following scheme:



Evidence for unimolecular hairpin structure arises from NMR experiments. The equilibrium between the duplex and the hairpin is slow on the NMR time scale, so that resonances from both species can be observed simultaneously under the appropriate conditions. The equilibrium constants can be determined from the ratio of the areas of equivalent peaks. Thermodynamic parameters for the duplex-hairpin equilibrium are estimated from a van't Hoff analysis of the equilibrium constants at several temperatures well below the temperature when unfolded conformations become populated⁵.

Experimental constraints

Quantitative distance constraints were obtained from NOESY experiments by using a complete relaxation matrix analysis with the program MARDIGRAS⁶. Error bounds in the inter-protonic distances were estimated by carrying out several MARDIGRAS calculations with different initial models, mixing times and correlation times. Standard A- and B-form duplexes and an average G-quadruplex structure were used as initial models, and three correlation times (1.0, 2.0 and 4.0 ns) were employed, assuming an isotropic motion for the molecule. Experimental intensities were recorded at three different mixing times (100, 150 and 250 ms) for non-exchangeable protons. Final constraints were obtained by averaging the upper and lower distance bounds in all the MARDIGRAS runs. Qualitative limits of 1.8 Å and 5 Å were set in those distances where no quantitative analysis could be carried out due to overlapping cross-peaks or peaks with very weak intensities. In addition to these experimentally derived constraints, Watson-Crick and Hoogsteen hydrogen bond restraints were used for duplex and quadruplex, respectively. In the quadruplex calculation planarity restraints were also imposed during structure calculations. Target values for distances and angles related to hydrogen bonds were set as described from crystallographic data. ¹⁹F-¹H distance constraints from qualitative analysis of HOESY experiments, in chapter 5, 6 and 7, were used in the calculations. No backbone angle constraints were applied. Distance constraints with their corresponding error bounds were incorporated into the AMBER potential energy by defining a flat-well potential term.

¹H-¹H *J*-coupling constants could not be accurately measured due to the relatively broad line-widths of the sugar proton signals. However, sum of *J*-coupling constants were roughly estimated from DQF-COSY cross-peaks and some ¹H-¹⁹F *J*-coupling constants could be measured using NOESY and DQF-COSY spectra. Loose values were set for the sugar dihedral angles δ , ν_1 and ν_2 to constrain them.

Duplex structure determination

Structures were calculated with the SANDER module of the molecular dynamics package AMBER 7.0⁷. Starting models of the duplexes were built in the A- and B-canonical structures using SYBYL. These structures were taken as starting points for the AMBER refinement, which started with a short run *in vacuo* (using hexahydrated Na⁺ counterions placed near the phosphates to neutralize the system). The resulting structures from *in vacuo* calculations were refined including explicit solvent, periodic boundary conditions and the Particle-Mesh-Ewald method to evaluate long-range electrostatic interactions⁸. Thus, the structures obtained in the previous step were placed in the center of a water-box with around 4000 water molecules and sodium counterions to obtain electroneutral systems. We used the parmbsc0⁹ revision of the parm99 force field^{10,11} including suitable parameters for the modified nucleotides. The TIP3P model was used to describe water molecules¹². The protocol for the molecular dynamics refinement consisted of an equilibration period of 160 ps using a standard equilibration process¹³, followed by 10 independent 500 ps runs. Final structures were obtained by averaging the last ps of individual trajectories and further relaxation of the structure. Analysis of the representative structures as well as the MD trajectories was carried out with the programs CURVES V5.1¹⁴, MOLMOL¹⁵, the analysis tools of AMBER and SYBYL, and additional “in house” programs.

G-quadruplex structure determination

Initial structure computations were performed using XPLOR-NIH program following three standard steps as previously described^{16,17}: (I) distance geometry simulated annealing, (II) distance-restrained molecular dynamics refinement, and (III) relaxation matrix intensity refinement. Dynamic calculations *in vacuo* were successively performed obtaining 10 best structures. Two XPLOR structures, with the highest RMSD value between them, were taken as starting points for the AMBER refinement, which was carried out following the same protocols described for duplexes.

Molecular modeling of the hairpin structure

Initial structural models of the hairpin species were built with the program SYBYL on the basis of qualitative NMR information. Conformation of 2'-F-ANA residues was set according to standard B-form parameters. Initial models were submitted to a molecular dynamic calculation following the same protocols described for duplexes. Only hydrogen bonds restraints for the Watson-Crick G-C base pairs in the stem were used.

BIBLIOGRAPHY

- (1) Damha, M. J.; Ogilvie, K. K. *Oligoribonucleotide Synthesis. The silyl-phosphoramidite method*. **Methods Mol. Biol.** (1993), 20, 81.
- (2) Piatto, M.; Saudek, V.; Sklenar, V. *Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions* **J. Biomol. NMR** (1992), 2, 661.
- (3) Yu, C.; Levy, G. C. *Two-dimensional heteronuclear NOE (HOESY) experiments: investigation of dipolar interactions between heteronuclei and nearby protons* **J. Am. Chem. Soc.** (2002), 106, 6533.
- (4) Goddard, D. T.; Kneller, G.; 3. ed.; University of California: San Francisco.
- (5) Marky, L. A.; Breslauer, K. J. *Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves* **Biopolymers** (1987), 26, 1601.
- (6) Borgias, B. A.; James, T. L. *MARDIGRAS, a procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure* **J. Magn. Reson.** (1990), 87, 475.
- (7) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; III, T. E. C.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A.; 5 ed.; University of California: San Francisco, 1997.
- (8) Darden, T. E.; York, D.; Pedersen, L. *Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems* **J. Chem. Phys.** (1993), 98, 10089.
- (9) Perez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T.; Laughton, C.; Orozco, M. *Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers* **Biophys. J.** (2007), 92, 3817.
- (10) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat* **J. Biomol. Struct. Dynamics** (1999), 16, 845.
- (11) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *A 2nd generation force field for the simulation of proteins, nucleic acids and organic molecules* **J. Am. Chem. Soc.** (1995), 117, 5179.
- (12) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *Comparison of simple potential functions for simulating liquid water* **J. Chem. Phys.** (1983), 79, 926.
- (13) Shields, G. C.; Laughton, C. A.; Orozco, M. *Molecular Dynamics Simulations of the d(T.A.T) Triple Helix* **J. Am. Chem. Soc.** (1997), 119, 7463.
- (14) Lavery, R.; Sklenar, H.; 3.0 ed.; Laboratory of Theoretical Biochemistry CNRS: Paris, 1990.
- (15) Koradi, R.; Billeter, M.; Wuthrich, K. *MOLMOL: a program for display and analysis of macromolecular structures* **J. Mol. Graph.** (1996), 14, 51.
- (16) Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. *Using Xplor-NIH for NMR Molecular Structure Determination* **ChemInform** (2006), 37, no.
- (17) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Marius Clore, G. *The Xplor-NIH NMR molecular structure determination package* **J. Magn. Reson.** (2003), 160, 65.

Annex II

Tables

CHAPTER 5 ANNEX

The solution structure of double helical arabino-nucleic acids (ANA and 2′F-ANA): effect of arabinoses in duplex-hairpin interconversion.

CONTENTS:

Table A5.1: Assignment of the proton resonances of **gap(FA)**.

Table A5.2 J couplings in Hz of **gap(FA)**.

Table A5.3. Some significant helical parameters of the duplex form of **gap(FA)** and comparison with calculated parameters of unmodified Tjandra's dodecamer (PDB: 1DUF).

Table A5.4. Some significant base-base helical parameters of the duplex form of **gap(FA)** and comparison with calculated parameters of unmodified Tjandra's dodecamer (PDB:1DUF).

Table A5.5. Average dihedral angles and order parameters of the duplex structure of **gap(FA)**.

Table A5.6. Average dihedral angles and order parameters of the Tjandra's dodecamer duplex structure of 5′-CGCGAATTCGCG-3′ (PDB:1DUF).

Table A5.1: Assignment of the proton resonances of **gap(FA)** (100 mM NaCl, 25 mM sodium phosphate, pH 7). Residues 1 to 4 and 9 to 12 are 2'-F-ANA, residues 5 to 8 are ANA.

Residue	NH	NH ₂ (2)	NH ₂ (1)	H6/H8	H2/H5/M	H1'	H2''	F	H3'	H4'
1 CYT	-	n.o.	8.2	7.5	5.8	5.8	5.2	-120.5	n.o.	3.9
2 GUA	13.0	n.o.	n.o.	7.5	-	5.7	5.4	-119.5	4.7	4.1
3 CYT	-	6.6	8.6	6.9	5.0	5.6	5.2	-118.5	4.8	3.9
4 GUA	12.5	7.6	n.o.	7.3	-	5.7	5.4	-119.2	4.8	4.0
5 ADE	-	n.o.	n.o.	7.5	7.0	5.8	n.o.	-	n.o.	n.o.
6 ADE	-	6.2	7.7	n.o.	7.4	5.8	4.4	-	n.o.	n.o.
7 URA	14.0	-	-	6.9	4.5	5.6	4.4	-	n.o.	4.0
8 URA	13.8	-	-	7.3	5.0	5.8	n.o.	-	n.o.	4.1
9 CYT	-	6.9	8.6	7.1	5.4	5.7	5.3	-118.8	n.o.	n.o.
10 GUA	12.8	n.o.	n.o.	7.3	-	5.4	5.7	-118.6	4.8	4.0
11 CYT	-	6.8	8.7	6.9	5.0	5.7	5.3	-117.8	n.o.	n.o.
12 GUA	n.o.	n.o.	n.o.	7.3	-	5.9	4.9	-118.6	4.4	3.8

n.o. not observed

Table A5.2: *J* couplings in Hz of **gap(FA)**. (100 mM NaCl, 25 mM sodium phosphate, T=25 °C, pH 7, D₂O). Residues 1 to 4 and 9 to 12 are 2'-F-ANA, residues 5 to 8 are ANA.

Residue	F-H1'	F-H2''	F-H3'	F-H4'	H2''-H1'	H2''-H3'	H3'-H4'	F-H8/H6
1 CYT	19	52	-	0	1-3	0	n.a	0
2 GUA	22	49	16	0	1-3	0	n.a	7
3 CYT	21	51	25	0	1-3	0	n.a	0
4 GUA	25	51	22	0	1-3	0	n.a	5
5 ADE	-	-	-	-	1-2	0	n.a	-
6 ADE	-	-	-	-	1-2	0	n.a	-
7 THY	-	-	-	-	1-2	0	n.a	-
8 URA	-	-	-	-	1-2	0	n.a	-
9 CYT	24	51	n.a	n.a	1-3	0	n.a	0
10 GUA	22	50	22	0	1-3	0	n.a	5
11 CYT	24	50	n.a	n.a	1-3	0	n.a	0
12 GUA	16	52	24	0	1-3	0	n.a	4

n.a. not assigned

Table A5.3: Some significant helical parameters of the duplex form of **gap(FA)** and comparison with calculated parameters of unmodified Tjandra's dodecamer (PDB: 1DUF).

FANA-ANA- FANA duplex	Shift Dx	Slide Dy	Rise Dz	Tilt t	Roll r	Twist W
C1/G2	0.2	-1.5	3.7	1	13	33
G2/C3	-0.1	-1.1	3.4	1	1	33
C3/G4	0.7	-0.9	3.0	5	0	38
G4/A5	0.4	-1.7	3.5	-3	1	31
A5/A6	-0.5	-1.0	3.4	1	-11	37
A6/U7	0.1	-1.7	3.4	-0.3	-3	33
U7/U8	0.1	-1.1	2.7	8	-9	38
U8/C9	0.3	-1.5	3.0	6	1	31
C9/G10	-0.1	-0.9	3.5	-2	11	38
G10/C11	0.3	-1.0	3.3	1	-2	32
C11/G12	0.4	-1.35	3.4	-2	0	34

Tjandra's duplex	Shift Dx	Slide Dy	Rise Dz	Tilt t	Roll r	Twist W
C1/G2	0.7	-0.2	4.1	-6	7	32
G2/C3	0.7	-0.8	3.2	0	-2	35
C3/G4	0.5	-0.2	3.4	0	7	32
G4/A5	0.2	-0.7	3.4	-2	4	36
A5/A6	0.3	-0.8	3.2	-1	-3	37
A6/U7	0.7	-1.0	3.4	0	-3	36
U7/U8	1.1	-0.5	3.2	1	-3	37
U8/C9	0.7	-0.4	3.4	2	4	36
C9/G10	0.03	-0.3	3.4	0	7	32
G10/C11	0.8	-0.7	3.2	-1	-2	35
C11/G12	-0.04	0.09	4.1	6	8	32

Table A5.4: Some significant base-base helical parameters of the duplex form of **gap(FA)** and comparison with calculated parameters of unmodified Tjandra's dodecamer (PDB:1DUF).

FANA-ANA-FANA duplex	Stretch (Sy)	Stagger (Sz)	Opening (Sigma)	Inclin (eta)
C1/G24	-0.00	-0.08	2.2	5.66
G2/C23	0.01	0.01	0.28	0.20
C3/G22	0.03	0.23	0.73	-5.22
G4/C21	-0.02	-0.17	2.15	-4.18
A5/U20	0.04	0.38	3.88	-8.08
A6/U19	0.04	0.85	1.60	-10.9
U7/A18	0.06	0.88	2.54	-10.9
U8/A17	0.05	0.38	3.92	-7.70
C9/G16	-0.01	-0.19	1.99	-2.70
G10/C15	0.02	0.20	0.43	-1.88
C11/G14	0.00	0.07	0.31	3.82
G12/C13	-0.04	-0.15	0.03	7.73

Tjandra's duplex	Stretch (Sy)	Stagger (Sz)	Opening (Sigma)	Inclin (eta)
C1/G24	-0.2	-0.6	2.0	12.3
G2/C23	-0.2	0.1	1.6	4.00
C3/G22	-0.3	-0.4	3.6	3.20
G4/C21	-0.2	-0.3	1.8	1.25
A5/T20	-0.2	-0.5	2.5	0.40
A6/T19	-0.2	-0.6	0.5	1.78
T7/A18	-0.2	-0.6	0.6	2.13
T8/A17	-0.2	-0.5	3.0	1.50
C9/G16	-0.2	-0.2	1.6	2.23
G10/C15	-0.3	-0.4	3.3	3.98
C11/G14	-0.2	0.1	1.8	4.53
G12/C13	-0.2	-0.6	2.1	12.53

Table A5.5: Average dihedral angles and order parameters of the duplex structure of *gap(FA)*. *P* means Phase Angle, *Am* means Amplitude, *A* means Average, and *O.P.* means Order Parameter.

	Pseudorot		α		β		γ		δ		ϵ		ζ		χ	
	P	Am	A	O.P.	A	O.P.	A	O.P.	A	O.P.	A	O.P.	A	O.P.	A	O.P.
1C	123	37	-	-	-	-	60	1	113	1	-161	1	-81	0.9	-141	1
2G	138	37	-70	1	-176	1	52	1	120	1	-176	1	-101	1	-115	1
3C	153	34	-70	1	175	1	58	1	130	1	-176	1	-97	1	-111	1
4G	143	35	-66	1	175	1	59	1	124	1	-177	1	-92	1	-115	1
5A	124	41	-71	1	163	1	59	1	121	0.9	-144	0.7	-124	1	-131	0.9
6A	142	35	-68	0.9	176	0.9	57	1	130	1	-174	1	-97	0.7	-124	1
7U	132	36	-64	1	167	1	57	1	122	1	-176	1	-98	1	-121	1
8U	121	35	-66	1	171	1	63	1	115	0.9	-173	1	-88	1	-127	1
9C	129	40	-75	1	-170	1	56	1	117	1	-166	0.9	-112	1	-120	1
10G	177	29	-66	0.9	172	0.9	52	1	135	1	-177	1	-90	0.9	-110	1
11C	125	32	-66	1	167	1	60	1	114	1	-174	1	-89	1	-122	1
12G	106	36	-71	0.9	-177	0.3	61	1	105	0.8	-	-	-	-	-137	0.9
13C	120	37	-	-	-	-	59	1	110	1	-163	1	-84	0.9	-143	1
14G	143	36	-69	1	-177	1	53	1	123	1	-176	1	-99	1	-114	1
15C	153	34	-71	1	175	1	57	1	129	1	-176	1	-96	1	-111	1
16G	141	35	-65	1	172	1	60	1	123	1	-177	1	-90	1	-116	1
17A	120	40	-68	1	171	1	59	1	117	0.9	-160	0.8	-113	1	-135	0.9
18A	155	34	-68	1	177	0.9	57	1	136	1	-175	1	-97	0.8	-119	1
19U	130	36	-64	1	166	1	56	1	120	0.9	-175	1	-97	1	-120	1
20U	120	36	-66	1	173	1	64	1	114	1	-174	1	-87	1	-129	1
21C	130	40	-74	1	-169	1	57	1	118	1	-167	0.9	-116	1	-119	1
22G	182	29	-67	1	170	0.9	53	1	137	1	-177	1	-90	0.9	-109	1
23C	124	34	-65	1	168	1	62	1	114	1	-175	1	-89	1	-122	1
24G	100	37	-77	0.9	-175	0.5	60	1	91	0.8	-	-	-	-	-141	0.9

Table A5.6: Average dihedral angles and order parameters of the Tjandra's dodecamer duplex structure of 5'-CGCGAATTCGCG-3' (PDB:1DUF). *P* means Phase Angle, *Am* means Amplitude, *A* means Average, and *O.P* means Order Parameter.

	Pseudorot		α		β		γ		δ		ϵ		ζ		χ	
	P	Am	A	O.P	A	O.P.	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P
1C	110	27	-	-	-	-	60	1	118	1	-170	1	-81	1	-126	1
2G	140	30	-73	1	-161	1	45	1	132	1	172	1	-99	1	-103	1
3C	115	29	-63	1	177	1	57	1	117	1	178	1	-85	1	-123	1
4G	155	35	-72	0.9	-167	0.9	50	1	138	1	173	1	-104	1	-102	1
5A	152	29	-57	1	-175	1	49	1	134	1	176	1	-96	1	-115	1
6A	149	29	-69	1	-172	1	55	1	131	1	175	1	-96	1	-106	1
7U	119	30	-57	1	175	0.9	53	1	117	1	177	1	-82	1	-120	1
8U	136	28	-59	1	-173	1	43	1	124	1	178	1	-90	1	-108	1
9C	116	32	-69	1	177	1	57	1	114	1	177	1	-84	1	-122	1
10G	143	29	-75	1	-164	0.9	48	0.9	128	1	177	1	-94	0.9	-105	1
11C	110	32	-53	1	170	0.9	52	1	112	1	-174	1	-84	1	-124	1
12G	113	28	-77	1	-177	1	58	1	113	1	-	-	-	-	-114	1
13C	110	27	-	-	-	-	60	1	118	1	-170	1	-81	1	-126	1
14G	140	30	-73	1	-161	1	45	1	132	1	172	1	-99	1	-103	1
15C	115	29	-63	1	177	1	57	1	117	1	178	1	-85	1	-123	1
16G	155	35	-72	0.9	-167	0.9	50	1	138	1	173	1	-104	1	-102	1
17A	152	29	-57	1	-175	1	49	1	134	1	176	1	-96	1	-115	1
18A	149	29	-69	1	-172	1	55	1	131	1	175	1	-96	1	-106	1
19U	119	30	-57	1	175	0.9	53	1	117	1	177	1	-82	1	-120	1
20U	136	28	-59	1	-173	1	43	1	124	1	178	1	-90	1	-108	1
21 C	116	32	-69	1	177	1	57	1	114	1	177	1	-84	1	-122	1
22G	143	29	-75	1	-164	0.9	48	0.9	128	1	177	1	-94	0.9	-105	1
23C	110	32	-53	1	170	0.9	52	1	112	1	-174	1	-84	1	-124	1
24G	113	28	-77	1	-177	1	58	1	113	1	-	-	-	-	-114	1

CHAPTER 6 ANNEX

**The solution structure of 2′F-ANA:RNA and ANA:RNA
hybrid duplexes. Roles of structure and hydrogen
bonding in their different stability**

CONTENTS:

Table A6.1: Assignment of the proton and fluorine resonances in **FR**.

Table A6.2: Assignment of the proton resonances in **AR**.

Table A6.3: Average values of the torsion angles in **FR**.

Table A6.4: Average values of the torsion angles in **AR**.

Table A6.5: Local inter-base helical parameters for **FR**.

Table A6.6: Local inter-base helical parameters for **AR**.

Table A6.1: Assignment of the proton and fluorine resonances of **FR** (0.8 mM oligonucleotide concentration, 25 mM sodium phosphate, 100mM NaCl, T=5 °C. pH 7). Residues 1 to 10 are 2'F-ANA, 11 to 20 are RNA.

Residue	NH	NH ₂ (2)	NH ₂ (1)	H6/H8	H2/H5/M	H1'	H2'/2''	F	H3'	H4'
1 GUA	13.44	n.o	n.o	7.98	-	6.17	5.67	-120.2	4.87	4.26
2 CYT	-	7.11	8.76	7.30	5.31	6.09	5.60	-119.5	4.80	4.40
3 THY	13.72	-	-	7.28	1.58	5.96	5.58	-117.2	4.88	4.19
4 ADE	-	7.98	6.28	7.88	7.13	6.28	5.66	-118.1	5.66	4.42
5 THY	13.32	-	-	6.96	1.45	5.74	5.40	-116.5	4.84	4.12
6 ADE	-	7.88	5.94	7.70	6.69	6.10	5.57	-117.8	5.04	4.29
7 ADE	-	7.99	6.50	7.65	7.60	6.14	5.56	-117.5	5.00	4.36
8 THY	13.48	-	-	6.80	1.34	5.77	5.46	-116.7	5.46	4.11
9 GUA	12.36	8.14	5.76	7.27	-	5.99	5.42	-119.4	4.89	4.28
10 GUA	n.o	n.o	n.o	7.22	-	6.14	5.06	-119.2	4.58	4.12
11 CYT	-	7.17	8.10	8.00	5.96	5.60	4.47	-	4.26	4.45
12 CYT	-	7.14	8.55	8.01	5.64	5.57	4.26	-	4.16	n.o
13 ADE	-	7.69	6.63	8.05	7.25	5.89	4.60	-	4.23	4.37
14 URA	13.99	-	-	7.82	5.01	5.48	4.37	-	4.46	4.63
15 URA	13.06	-	-	8.12	5.63	5.71	4.46	-	4.34	4.23
16 ADE	-	n.a	6.78	8.17	7.21	5.91	4.37	-	4.20	4.52
17 URA	13.15	-	-	7.71	5.04	5.43	4.56	-	4.44	4.51
18 ADE	-	7.78	6.48	8.06	6.77	5.92	4.24	-	4.41	4.63
19 GUA	13.27	n.o	6.09	7.28	-	5.57	4.46	-	n.a	n.a
20 CYT	-	7.05	8.28	7.51	5.27	5.78	4.00	-	4.15	7.51

n.o not observed.

n.a not assigned.

Table A6.2: Assignment of the proton resonances of **AR** (0.8 mM oligonucleotide concentration, 25 mM sodium phosphate, 100 mM NaCl, T=5 °C, pH 7). Residues 1 to 10 are ANA, 11 to 20 are RNA.

Residue	NH	NH ₂ (2)	NH ₂ (1)	H6/H8	H2/H5/M	H1'	H2'/2''	H3'	H4'
1 GUA	13.41	n.o	n.o	8.13	-	6.10	4.74	n.a	3.93
2 CYT	-	6.91	8.68	7.41	5.31	6.10	4.69	4.40	4.26
3 URA	13.59	-	-	7.54	5.23	5.98	4.77	4.26	4.37
4 ADE	-	6.30	7.84	8.12	7.09	6.14	4.72	4.20	4.32
5 URA	13.24	-	-	7.16	4.83	7.73	4.61	4.41	4.20
6 ADE	-	6.30	7.84	7.93	6.60	5.98	4.74	4.21	4.26
7 ADE	-	6.54	7.86	7.65	7.57	5.98	4.59	4.27	n.a
8 URA	13.40	-	-	7.01	4.63	5.76	4.63	4.36	4.15
9 GUA	12.29	5.62	8.02	7.56	-	5.91	4.44	4.17	4.57
10 GUA	n.o	n.o	n.o	7.33	-	6.11	4.39	4.10	4.29
11 CYT	-	8.15	7.12	8.09	5.95	5.50	4.51	4.26	4.43
12 CYT	-	8.51	7.10	8.02	5.65	5.59	4.78	4.63	4.44
13 ADE	-	7.82	6.50	8.06	7.27	5.92	4.55	4.61	4.25
14 URA	14.11	-	-	7.71	5.60	5.45	4.33	4.55	4.44
15 URA	13.24	-	-	8.09	5.59	5.66	4.69	4.10	4.62
16 ADE	-	5.96	7.76	8.15	6.80	5.92	4.36	4.62	4.54
17 URA	13.27	-	-	7.69	5.07	5.40	4.51	4.10	4.39
18 ADE	-	7.85	6.42	8.04	6.81	5.97	4.63	4.61	4.52
19 GUA	13.41	6.02	8.82	7.24	-	5.54	4.40	4.03	4.23
20 CYT	-	8.32	7.04	7.47	5.24	5.76	4.00	4.22	4.15

n.o not observed.

n.a not assigned.

Table A6.3: Average values of the torsion angles in **FR**. *P* means Phase Angle, *Am* means Amplitude, *A* means Average, and *O.P* means Ordet Parameter. Residues 1 to 10 are 2'F-ANA, 11 to 20 are RNA.

	Pseudorot		α		β		γ		δ		ϵ		ζ		χ	
	P	Am	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P.	A	O.P	A	O.P.
1G	153	40	-	-	-	-	46	0.6	132	1	-170	1	-94	0.8	-111	1
2C	133	36	-70	1	176	1	52	1	116	1	-171	1	-95	1	-125	1
3T	127	35	-73	0.9	161	0.7	73	0.7	115	1	-162	1	-87	1	-132	1
4A	149	34	-75	1	178	1	45	1	124	1	178	1	-80	1	-105	1
5T	96	82	-65	1	166	1	63	1	97	1	-165	1	-91	1	-141	1
6A	116	37	-74	1	177	1	54	1	103	1	-172	1	-76	1	-133	1
7A	129	38	-65	1	174	1	55	1	113	1	-175	1	-87	1	-117	1
8T	94	74	-67	1	168	1	61	1	97	1	-168	1	91	1	-137	1
9G	155	39	-118	0.4	163	0.7	116	0.4	134	1	-180	0.8	-73	0.9	-131	1
10G	167	35	-73	1	-162	1	39	0.9	137	1	-	-	-	-	-115	1
11C	36	38	-	-	-	-	62	0.7	77	1	-164	1	-68	1	-151	1
12C	23	29	-85	1	180	1	65	1	79	1	-155	1	-69	1	-158	1
13A	47	43	-76	1	171	1	70	1	85	1	-173	1	-64	1	-157	1
14U	96	28	-85	1	-170	1	66	1	106	0.9	-170	1	-78	1	-133	0.9
15U	35	44	-75	1	165	1	62	1	76	1	-166	1	-74	1	-160	1
16A	32	37	-85	0.8	173	1	77	0.9	80	1	-165	1	-69	1	-164	1
17U	18	27	-86	1	180	1	67	1	81	1	-159	1	-76	1	-157	1
18A	26	29	-78	1	170	1	72	1	83	1	-165	1	-65	1	-163	1
19G	18	29	-95	0.6	176	1	85	0.7	80	1	-158	1	-73	1	-158	1
20C	34	33	-76	1	177	1	65	1	90	1	-	-	-	-	-155	1

Table A6.4: Average values of the torsion angles in **AR**. *P* means Phase Angle, *Am* means Amplitude, *A* means Average, and *O.P* means Order Parameter. Residues 1 to 10 are ANA, 11 to 20 are RNA

	Pseudorot		α		β		γ		δ		ϵ		ζ		χ	
	P	Am	A	O.P	A	O.P	A	O.P	A	O.P.	A	O.P.	A	O.P	A	O.P.
1G	155	36	-	-	-	-	34	0.6	135	0.9	-170	0.9	-109	1	-109	1
2C	104	38	-70	1	166	1	59	1	105	1	-171	1	-96	1	-136	1
3U	131	34	-72	1	-178	1	54	1	121	1	-169	1	-88	1	-130	1
4A	142	43	-70	1	173	1	54	1	132	1	-169	0.9	-89	1	-115	1
5U	158	33	-63	1	175	1	56	1	137	1	-172	1	-127	0.9	-122	1
6A	139	39	-71	1	178	1	53	1	126	1	-175	1	-92	1	-119	1
7A	127	33	-68	1	168	0.8	64	0.8	117	1	-174	1	-95	1	-122	1
8U	116	42	-66	1	169	1	54	1	109	1	-172	1	-87	1	-137	1
9G	150	40	-69	1	-179	1	48	1	136	1	-173	1	-86	1	-117	1
10G	79	27	78	-	170	1	54	0.9	98	0.8	-	-	-	-	-135	1
11C	8.1	30	-	-	-	-	33	0.6	92	1	-139	1	-72	1	-166	1
12C	15	36	-80	1	170	1	61	1	85	1	-158	1	-72	1	-156	1
13A	12	36	-81	1	174	1	64	1	85	1	-162	1	-70	1	-158	1
14U	28	35	-76	1	169	1	72	1	83	1	-167	1	-68	1	-158	1
15U	45	41	-87	0.8	173	1	78	0.9	83	1	-164	1	-73	1	-153	1
16A	14	38	-78	1	169	1	69	1	83	1	-165	1	-64	1	-163	1
17U	44	37	-83	0.8	172	1	78	1	85	1	-160	1	-69	1	-156	1
18A	27	39	-84	0.8	173	1	75	0.9	84	1	-170	1	-68	1	-158	1
19G	39	44	-82	1	173	1	66	1	73	1	-157	1	-59	1	-171	1
20C	21	42	-79	1	178	1	64	1	79	1	-	-	-	-	-163	1

Tables A6.5: Local inter-base helical parameters for **FR**.

2'F-ANA strand	Shift Dx	Slide Dy	Rise Dz	Tilt t	Roll r	Twist W
G1/C2	-0.3	-1.2	3.6	-4.1	8.5	29.4
C2/T3	0.7	-1.2	3.2	4.2	-3.8	37.0
T3/A4	-0.1	-0.9	3.4	1.7	16.0	28.7
A4/T5	1.3	-1.3	3.2	1.8	-0.9	33.9
T5/A6	-0.3	-2.0	3.9	-9.8	17.6	28.7
A6/A7	-0.2	-1.3	3.0	9.7	-0.42	29.2
A7/T8	1.0	-1.5	3.2	3.8	1.10	32.7
T8/G9	0.1	-1.6	3.7	-3.2	19.3	27.1
G9/G10	0.4	-2.3	3.6	-6.6	3.3	39.3

RNA strand	Shift Dx	Slide Dy	Rise Dz	Tilt t	Roll r	Twist W
G19/C20	-0.8	-1.8	3.4	-3.4	12.5	30.2
A18/G19	0.2	-1.8	3.0	-11.2	-4.3	32.0
U17/A18	-1.3	-1.8	3.2	-4.0	8.1	34.5
A16/U17	0.9	-1.9	3.1	-9.2	-6.9	23.8
U15/A16	-1.0	-2.2	3.5	0.3	10.4	33.2
U14/U15	-1.2	-1.5	3.0	-10.3	3.1	32.2
A13/U14	0.5	-1.8	3.1	-6.9	-0.3	27.2
C12/A13	-1.1	-1.8	3.5 5	-3.6 8	13.3	29.5
C11/C12	-0.5	-1.5	4.2	3.8	15.0	36.9

Tables A6.6: Local inter-base helical parameters for **AR**.

ANA strand	Shift Dx	Slide Dy	Rise Dz	Tilt t	Roll r	Twist W
G1/C2	0.8	-1.4	3.6	-4.1	3.3	34.4
C2/U3	-0.5	-1.9	3.9	-7.5	11.9	29.3
U3/A4	0.6	-1.5	3.1	3.7	3.6	29.7
A4/U5	-0.02	-1.8	3.3	-2.2	-0.9	37.9
U5/A6	-0.04	-1.7	2.9	1.9	1.6	34.2
A6/A7	0.4	-1.4	2.9	7.4	-10.5	34.7
A7/U8	0.3	-1.9	3.2	-2.7	-0.2	28.4
U8/G9	0.8	-1.9	3.2	-0.3	7.2	34.9
G9/G10	-0.08	-2.1	3.2	0.2	0.9	37.2

RNA strand	Shift Dx	Slide Dy	Rise Dz	Tilt t	Roll r	Twist W
G19/C20	-0.09	-2.7	3.4	0.1	8.1	30.2
A18/G19	-1.0	-2.3	3.2	2.3	0.8	35.1
U17/A18	-0.5	-1.5	3.2	-2.6	13.4	30.0
A16/U17	-0.3	-1.6	2.9	-10.3	-4.9	32.8
U15/A16	-0.8	-1.5	3.2	-2.8	0.5	35.7
U14/U15	-0.5	-1.6	3.0	-9.3	-0.2	32.4
A13/U14	-0.09	-1.9	3.1	-6.8	-3.4	27.1
C12/A13	-0.3	-1.6	3.4	-3.3	3.3	34.6
C11/C12	-0.6	-1.6	3.4	-3.8	5.7	38.1

CHAPTER 7 ANNEX

Stabilizing FC-H...O electrostatic interactions in 2'-substituted Nucleic Acids

CONTENTS:

Table A7.1: Assignment of the proton and fluorine resonances of the **H** duplex.

Table A7.2: Assignment of the proton and fluorine resonances of **S** duplex.

Table A7.3: Average dihedral angles and order parameters of the structure of **H** duplex.

Table A7.4: Average dihedral angles and order parameters of the structure of **S** duplex.

Table A7.5: Distances $H2'/H2''_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2'/H2''_{i..}O4'_{i+1}$ in **H** duplex.

Table A7.6: Distances $H2'/H2''_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2'/H2''_{i..}O4'_{i+1}$ in **S** duplex.

Table A7.7: Distances $H2'_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2'_{i..}O4'_{i+1}$ in full 2'-F-RNA (PDB ID: 3P4A) duplex.

Table A7.8: Distances $H2'_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2'_{i..}O4'_{i+1}$ in F-RNA-RNA (PDB ID: 3P4B) chimeric duplex.

Table A7.9: Distances $H2'_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2'_{i..}O4'_{i+1}$ in full RNA (PDB ID: 1RXA) duplex.

Table A7.10: Distances $H2''_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2''_{i..}O4'_{i+1}$ in full in FANA:RNA (PDB ID: 2KP4) duplex.

Table A7.1: Assignment of the proton and fluorine resonances of the **H** duplex. (Buffer conditions: 25 mM sodium phosphate, 100 mM NaCl, T=10 °C exchangeable and T=25 °C non exchangeable protons, pH 7). rf=2'F-RNA and af=2'F-ANA.

Residue	NH	NH ₂ (2)	NH ₂ (1)	H6/H8	H2/H5/M	H1'	H2'/H2''	F	H3'	H4'
1 rfC	-	8.37	6.99	8.01	5.95	5.73	5.35	-123.0	4.60	4.32
2 rfG	13.0	n.o.	n.o.	7.72	-	6.05	5.29	-123.3	4.83	4.55
3 rfC	-	8.59	6.86	7.52	5.33	5.77	5.39	-122.9	4.62	4.50
4 rfG	12.0	8.08	5.52	7.53	-	6.00	5.41	-123.2	4.72	4.55
5 rfA	-	7.87	6.87	7.61	7.24	6.18	5.41	-123.0	4.68	4.61
6 rfA	-	8.04	7.01	7.39	7.92	6.16	5.25	-121.6	4.51	4.55
7 afT	14.2	-	-	7.01	1.19	6.10	5.48	-115.9	4.72	4.34
8 afT	13.6	-	-	7.28	1.42	6.06	5.53	-117.4	4.82	4.19
9 afC	-	8.24	6.77	7.22	5.40	5.95	5.42	-117.0	4.86	4.15
10 afG	12.6	n.o.	n.o.	7.34	-	5.96	5.49	-119.3	4.96	4.25
11 afC	-	8.42	6.72	7.04	5.12	5.90	5.42	-118.0	4.76	4.19
12 afG	12.9	n.o.	n.o.	7.38	-	6.13	5.05	-119.1	4.59	4.05

n.o. not observed

Table A7.2: Assignment of the proton and fluorine resonances of **S** duplex. (Buffer conditions: 25 mM sodium phosphate, 100mM NaCl, T=10 °C exchangeable and T=25 °C non exchangeable protons, pH 7). rf=2'F-RNA and af=2'F-ANA.

Residue	NH	NH ₂ (2)	NH ₂ (1)	H6/H8	H2/H5/M	H1'	H2'/H2''	F	H3'	H4'
1 rfC	-	8.36	7.03	8.00	5.95	5.75	5.38	-123.3	4.61	4.35
2 rfG	13.2	n.o.	n.o.	7.72	-	6.09	5.31	-123.2	4.80	4.57
3 rfC	-	8.52	6.94	7.48	5.37	5.78	5.36	-122.8	4.62	4.52
4 rfG	12.0	7.84	5.53	7.52	-	6.04	5.35	-122.9	4.75	4.56
5 rfA	-	7.71	6.94	7.59	7.32	6.21	5.36	-122.9	4.73	4.61
6 rfA	-	7.89	7.07	7.36	7.88	6.15	5.24	-121.8	4.35	4.55
7 afT	14.1	-	-	6.90	1.28	6.00	5.42	-113.7	4.68	4.36
8 rfU	13.8	-	-	7.52	5.38	5.99	5.23	-112.8	n.o.	4.49
9 afC	-	8.30	6.90	7.43	5.51	5.98	5.40	-113.5	4.76	4.41
10 rfG	12.6	n.o.	n.o.	7.49	-	5.91	5.39	-125.9	n.o.	4.51
11 afC	-	8.48	6.78	7.19	5.12	6.00	5.30	-116.0	4.76	4.37
12 rfG	12.8	n.o.	n.o.	7.48	-	6.12	5.02	-126.2	4.53	4.26

n.o. not observed

Table A7.3: Average dihedral angles and order parameters of the structure of *H* duplex. rf=2'F-RNA and af=2'F-ANA.

	Pseudorot		α		β		γ		δ		ϵ		ζ		χ	
	P	Am	A	O.P	A	O.P	A	O.P	A	O.P.	A	O.P.	A	O.P	A	O.P.
1 rfC	63	46	-	-	-	-	62	0.9	75	1	-157	1	-69	1	-149	0.9
2 rfG	58	43	-79	1	171	1	62	1	79	1	-166	1	-65	1	-147	0.9
3 rfC	62	46	-76	1	174	1	63	1	74	1	-171	1	-82	1	-154	1
4 rfG	49	37	-76	1	173	1	64	1	84	1	-167	1	-70	1	-163	1
5 rfA	48	38	-74	1	169	1	69	1	81	1	-163	1	-65	1	-164	1
6 rfA	52	43	-69	1	175	1	68	1	76	1	-167	1	-71	1	-168	1
7 afT	137	30	-77	1	172	1	73	1	119	1	-158	1	-78	1	-150	1
8 afT	83	40	-75	1	177	1	55	1	92	1	-168	1	-78	1	-143	1
9 afC	105	35	-75	1	177	1	57	1	102	1	-168	1	-82	1	-136	1
10 afG	116	31	-68	1	176	1	54	1	108	0.9	-175	1	-90	1	-126	1
11 afC	143	39	-71	1	179	1	58	1	126	1	-172	1	-88	1	-125	0.9
12 afG	137	40	-82	0.9	171	0.7	53	0.9	124	0.9	-	-	-	-	-126	0.9
13 rfC	65	47	-	-	-	-	63	0.9	74	1	-157	1	-67	1	-151	0.9
14 rfG	57	43	-81	1	173	1	61	1	78	1	-163	1	-63	1	-153	0.9
15 rfC	60	43	-78	1	174	1	61	1	78	1	-171	1	-77	1	-153	1
16 rfG	51	37	-75	1	173	1	63	1	84	1	-166	1	-69	1	-161	1
17 rfA	50	40	-74	1	168	1	65	1	79	1	-163	1	-68	1	-164	1
18 rfA	54	43	-71	1	177	1	69	1	78	1	-168	1	-70	1	-168	1
19 afT	138	31	-75	1	174	1	72	1	120	1	-161	1	-79	1	-148	1
20 afT	85	39	-74	1	177	1	54	1	94	1	-168	1	-80	1	-146	1
21 afC	116	34	-72	1	172	1	57	1	108	1	-168	1	-82	1	-132	0.9
22 afG	85	37	-95	1	179	1	56	1	95	1	-164	1	-67	1	-131	0.9
23 afC	75	41	-63	0.9	174	1	36	1	88	1	-177	1	-81	1	-142	0.9
24 afG	150	35	-80	1	175	0.1	64	0.9	129	1	-	-	-	-	-130	0.9

Table A7.4: Average dihedral angles and order parameters of the structure of the *S* duplex. rf=2'F-RNA and af=2'F-ANA.

	Pseudorot		α		β		γ		δ		ϵ		ζ		χ	
	P	Am	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P
1 rfC	64	49	-	-	-	-	62	0.9	70	1	-162	0.9	-71	0.8	-152	0.9
2 rfG	59	45	-76	0.9	173	0.9	55	0.9	75	1	-168	0.9	-68	1	-151	0.9
3 rfC	37	39	-66	0.9	167	1	58	1	77	1	-168	0.9	-72	0.9	-151	1
4 rfG	56	38	-69	0.9	172	1	67	1	82	0.9	-162	0.9	-70	1	-169	1
5 rfA	52	40	-66	0.9	165	1	64	1	79	1	-164	1	-66	0.9	-167	1
6 rfA	55	44	-67	0.9	178	1	70	1	74	1	-165	0.9	-70	1	-172	1
7 afT	141	31	-70	1	166	1	69	0.9	119	0.9	-165	0.9	-83	1	-143	1
8 rfU	82	45	-68	0.9	179	1	56	1	81	1	-165	1	-79	0.9	-149	1
9 afC	150	32	-65	1	165	1	57	0.9	122	0.9	-170	1	-82	1	-126	1
10 rfG	87	46	-72	0.9	176	1	55	1	82	1	-165	1	-82	0.9	-146	1
11 afC	161	42	-129	1	154	1	57	1	138	1	-173	0.9	-75	1	-120	0.9
12 rfG	149	51	-76	0.9	176	0.7	63	0.9	37	0.9	-	-	-	-	-86	0.9
13 rfC	64	48	-	-	-	-	64	0.9	71	1	-159	1	-72	0.8	-154	0.9
14 rfG	58	45	-75	0.9	174	0.9	57	1	75	1	-167	0.9	-69	1	-152	0.9
15 rfC	40	38	-71	0.9	170	1	59	1	78	1	-165	0.9	-70	1	-150	1
16 rfG	59	39	-73	0.9	172	1	68	1	82	1	-163	0.9	-68	1	-167	1
17 rfA	54	42	-66	0.9	163	1	64	1	76	1	-161	0.9	-69	1	-165	1
18 rfA	57	45	-70	0.9	169	1	71	0.9	75	1	-165	1	-66	0.9	-170	1
19 afT	107	29	-70	1	171	1	75	0.9	105	0.9	-167	0.9	-77	1	-155	1
20 rfU	97	39	-71	0.9	177	1	60	1	97	0.9	-171	0.9	-80	1	-150	1
21 afC	136	31	-69	0.9	166	1	56	0.9	115	0.9	-167	0.9	-80	0.9	-132	0.9
22 rfG	91	46	-74	0.9	176	0.9	55	0.9	88	0.9	-167	0.9	-82	0.9	-142	0.9
23 afC	162	44	-143	0.9	152	1	57	0.9	139	1	-174	1	-75	0.9	-119	0.9
24 rfG	148	50	-78	0.9	176	0.8	64	1	25	0.9	-	-	-	-	-81	1

Table A7.5: Distances $H2'/H2''_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2'/H2''_{i..}O4'_{i+1}$ in **H** duplex. *rf*=2'F-RNA and *af*=2'F-ANA.

Nucleotides	Distance $H2'/H2''_{i..}O4'_{i+1}$	Angle $C2'-H2'/H2''_{i..}O4'_{i+1}$
rfC1-rfG2	3.0 ± 0.3	104
rfG2-rfC3	2.7 ± 0.1	108
rfC3-rfG4	3.8 ± 0.2	83
rfG4-rfA5	2.9 ± 0.3	106
rfA5-rfA6	2.7 ± 0.1	114
rfA6-afT7	3.1 ± 0.2	107
afT7-afT8	2.6 ± 0.1	166
afT8-afC9	2.7 ± 0.2	151
afC9-afG10	2.8 ± 0.2	167
afG10-afC11	3.1 ± 0.3	153
afC11-afG12	3.2 ± 0.2	139

Table A7.6: Distances $H2'/H2''_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2'/H2''_{i..}O4'_{i+1}$ in **S** duplex. *rf*=2'F-RNA and *af*=2'F-ANA.

Nucleotides	Distance $H2'/H2''_{i..}O4'_{i+1}$	Angle $C2'-H2'/H2''_{i..}O4'_{i+1}$
rfC1-rfG2	3.2 ± 0.2	99
rfG2-rfC3	3.2 ± 0.5	99
rfC3-rfG4	2.8 ± 0.1	110
rfG4-rfA5	3.0 ± 0.1	108
rfA5-rfA6	2.6 ± 0.1	114
rfA6-afT7	3.2 ± 0.1	106
afT7-rfU8	2.5 ± 0.1	159
rfU8-afC9	4.2 ± 0.1	77
afC9-rfG10	2.5 ± 0.1	143
rfG10-afC11	4.4 ± 0.2	71
afC11-rfG12	3.5 ± 0.2	140

Table A7.7: Distances $H2'_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2'_{i..}O4'_{i+1}$ in full 2'F-RNA (PDB ID: 3P4A) duplex. $rf=2'F$ -RNA.

Nucleotides	Distance $H2'_{i..}O4'_{i+1}$	Angle $C2'-H2'_{i..}O4'_{i+1}$
rfC1-rfG2	2.7	136
rfG2-rfA3	2.4	146
rfA3-rfA4	2.4	146
rfA4-rfT5	2.3	147
rfT5-rfT6	2.9	126
rfT6-rfC7	2.7	143
rfC7-rfG8	2.4	153

Table A7.8: Distances $H2'_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2'_{i..}O4'_{i+1}$ in F-RNA-RNA (PDB ID: 3P4B) chimeric duplex. $r=RNA$ y $rf=2'F$ -RNA.

Nucleotides	Distance $H2'_{i..}O4'_{i+1}$	Angle $C2'-H2'_{i..}O4'_{i+1}$
rfC1-rG2	2.7	137
rG2-rfA3	2.5	141
rfA3-rA4	2.5	136
rA4-rfT5	2.4	151
rfT5-rT6	3.1	123
rT6-rfC7	2.9	146
rfC7-rG8	2.4	151

Table A7.9: Distances $H2'_{i..}O4'_{i+1}$ (Å) and angles between $C2'-H2'_{i..}O4'_{i+1}$ in full RNA (PDB ID: 1RXA) duplex. $r=RNA$.

Nucleotides	Distance $H2'_{i..}O4'_{i+1}$	Angle $C2'-H2'_{i..}O4'_{i+1}$
rC1-rC2	2.4	145
rC2-rC3	3.2	144
rC3-rC4	2.2	156
rC4-rG5	2.7	138
rG5-rG6	2.1	144
rG6-rG7	2.3	140
rG7-rG8	3.5	132

Table A7.10: Distances $H2''_i \dots O4'_{i+1}$ (Å) and angles between $C2'-H2''_i \dots O4'_{i+1}$ in full in FANA:RNA (PDB ID: 2KP4) duplex. af=2'F-ANA.

Nucleotides	Distance $H2''_i \dots O4'_{i+1}$	Angle $C2'-H2''_i \dots O4'_{i+1}$
afG1-afC2	3.7 ± 0.3	127
afC2-afT3	3.0 ± 0.6	138
afT3-afA4	2.9 ± 0.2	163
afA4-afT5	3.0 ± 0.3	122
afT5-afA6	2.7 ± 0.2	159
afA6-afA7	2.9 ± 0.2	151
afA7-afT8	2.9 ± 0.2	135
afT8-afG9	2.7 ± 0.3	158
afG9-afG10	4.1 ± 0.5	134

CHAPTER 8 ANNEX

Effect of incorporation of 2'-deoxy-2'-difluorodeoxycytidine (gemcitabine) on the structure and stability of DNA:RNA hybrid duplexes

CONTENTS:

Table A8.1: Assignment of the proton and fluorine resonances of **[GEM]** hybrid duplex.

Table A8.2: Assignment of the proton and fluorine resonances of **[CON]** hybrid duplex.

Table A8.3: Average dihedral angles and order parameters of the structure of **[GEM]** duplex.

Table A8.4: Average dihedral angles and order parameters of the structure of **[CON]** duplex.

Table A8.5: Global base-pair axis parameters for **[GEM]** structure.

Table A8.6: Global base-pair axis parameters for **[CON]** structure.

Table A8.7: Global base-base buckle parameters for **[GEM]** and **[CON]** structures.

Table A8.8: Global inter-base tilt parameters for **[GEM]** and **[CON]** structures.

Table A8.1: Assignment of the proton and fluorine resonances of **[GEM]** hybrid duplex (Buffer conditions: 25 mM sodium phosphate, 100 mM NaCl, T=5 °C, pH 7).

Residue	NH	NH ₂ (2)	NH ₂ (1)	H6/H8	H2/H5/M	H1'	H2'/H2''	F	H3'	H4'
1 CYT	-	8.34	7.20	7.97	5.96	5.72	2.52/2.64	-	4.62	n.o.
2 THY	13.9	-	-	7.87	1.72	6.88	2.42/2.77	-	5.00	4.35
3 ADE	-	n.o.	6.20	8.32	7.14	6.22	2.70/2.86	-	5.03	4.20
4 dFdC	-	8.07	7.13	7.16	5.39	5.65	-	-35.3 -36.9	4.58	4.26
5 ADE	-	8.09	5.77	8.09	7.61	6.17	2.73/2.84	-	5.00	4.46
6 CYT	-	8.07	6.52	7.23	5.23	5.76	2.14/2.46	-	4.78	4.18
7 GUA	12.7	8.18	5.56	7.59	-	5.94	2.56/2.75	-	4.92	4.38
8 GUA	12.7	n.o.	n.o.	7.54	-	6.15	2.41	-	4.62	4.23
9 CYT	-	8.38	6.87	7.93	5.57	5.31	4.51	-	4.37	4.23
10 CYT	-	8.18	6.89	7.96	5.79	5.69	4.51	-	n.o.	n.o.
11 GUA	12.9	7.82	5.55	7.67	-	5.50	4.71	-	n.o.	4.42
12 URA	13.0	-	-	7.77	5.66	5.51	4.52	-	n.o.	n.o.
13 GUA	12.9	7.27	5.41	7.72	-	5.75	4.46	-	n.o.	n.o.
14 URA	13.2	-	-	7.77	5.14	5.41	4.60	-	n.o.	4.44
15 ADE	-	8.00	6.53	8.12	7.11	6.00	4.59	-	4.72	4.48
16 GUA	13.4	n.o.	n.o.	7.45	-	5.77	3.92	-	4.26	4.24

n.o. not observed

Table A8.2: Assignment of the proton and fluorine resonances of **[CON]** hybrid duplex (Buffer conditions: 25 mM sodium phosphate, 100 mM NaCl, $T=5\text{ }^{\circ}\text{C}$, pH 7).

Residue	NH	NH ₂ (2)	NH ₂ (1)	H6/H8	H2/H5/M	H1'	H2'/H2''	H3'	H4'
1 CYT	-	8.33	7.29	8.04	6.04	5.80	2.48/2.67	4.66	4.18
2 THY	13.93	-	-	7.88	1.76	6.15	2.48/2.80	4.99	4.37
3 ADE	-	8.02	6.23	8.28	7.15	6.21	2.67/2.85	5.02	4.44
4 CYT	-	8.24	6.83	7.36	5.28	5.75	2.18/2.48	4.81	4.22
5 ADE	-	n.o.	n.o.	8.08	7.52	6.21	2.61/2.84	4.97	4.38
6 CYT	-	8.18	6.57	7.25	5.17	5.77	2.14/2.44	4.80	4.18
7 GUA	12.73	n.o.	n.o.	7.60	-	5.93	2.56/2.73	4.91	4.36
8 GUA	n.o.	n.o.	n.o.	7.58	-	6.19	2.39	4.62	4.25
9 CYT	-	8.13	7.03	8.05	5.92	5.44	4.42	4.32	n.o.
10 CYT	-	8.42	6.92	7.92	5.61	n.o.	4.52	4.61	n.o.
11 GUA	12.97	n.o.	n.o.	7.70	-	5.69	4.55	4.70	n.o.
12 URA	13.43	-	-	7.75	5.16	5.48	4.53	4.73	n.o.
13 GUA	12.60	7.62	5.55	7.61	-	7.78	4.51	4.77	n.o.
14 URA	13.47	-	-	7.83	5.20	5.41	4.57	4.53	n.o.
15 ADE	-	7.99	6.62	8.11	7.07	6.00	4.42	4.72	n.o.
16 GUA	n.o.	n.o.	n.o.	7.44	-	5.77	4.26	n.o.	n.o.

n.o. not observed

Table A8.3: Average dihedral angles and order parameters of the structure of [GEM] duplex.

	Pseudorot		α		β		γ		δ		ϵ		ζ		χ	
	P	Am	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P
1 C	134	40	-	-	-	-	57	1	124	0.9	-179	0.9	-90	1	-134	0.9
2 T	125	35	-69	0.9	-179	0.9	51	1	117	0.9	-174	0.9	-91	0.9	-134	1
3 A	136	37	-68	0.9	179	0.9	49	1	124	0.9	-174	0.9	-92	0.9	-121	0.9
4dFdC	141	44	-67	0.9	172	0.9	51	1	126	1	-169	0.9	-98	0.9	-125	0.9
5 A	120	39	-62	0.9	168	0.9	51	1	112	0.9	-177	0.9	-94	1	-118	0.9
6 C	102	31	-68	0.9	174	0.9	57	0.9	102	0.9	-178	0.9	-91	0.9	-133	1
7 G	146	39	-68	0.9	179	1	56	1	133	1	-178	0.9	-94	0.9	-119	1
8 G	132	37	-93	1	178	0.9	54	0.9	124	0.9	-	-	-	-	-129	0.9
9 C	37	40	-	-	-	-	60	0.9	76	1	-174	0.9	-72	1	-145	0.9
10 C	18	43	-70	0.9	171	0.9	62	1	76	1	-163	0.9	-69	0.9	-153	1
11 G	6	41	-72	0.9	176	0.9	63	1	81	0.9	-157	0.9	-63	0.9	-160	1
12 U	15	38	-70	0.9	173	1	57	1	77	0.9	-169	0.9	-77	1	-163	1
13 G	9	37	-75	0.9	171	0.9	63	1	86	0.9	-154	0.9	-63	1	-143	0.9
14 U	13	39	-75	0.9	175	0.9	61	1	82	1	-161	0.9	-67	0.9	-159	0.9
15 A	35	23	-81	0.9	-178	0.9	63	1	94	0.9	-175	0.9	-81	1	-156	0.9
16 G	15	37	-80	0.9	-178	0.9	62	1	83	0.9	-	-	-	-	-163	0.9

Table A8.4: Average dihedral angles and order parameters of the structure of [CON] duplex.

	Pseudorot		α		β		γ		δ		ϵ		ζ		χ	
	P	Am	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P	A	O.P
1 C	182	17	-	-	-	-	43	1	126	0.9	-156	0.9	-76	1	-141	1
2 T	101	32	-67	0.9	-173	0.9	54	1	105	0.9	-179	0.9	-91	1	-130	1
3 A	158	25	-63	0.9	166	1	52	1	127	1	-175	1	-83	1	-123	1
4 C	100	34	-73	0.9	-176	0.9	52	1	100	1	-172	0.9	-89	1	-134	1
5 A	135	21	-68	0.9	167	0.9	50	1	118	0.9	-165	0.9	-86	1	-129	1
6 C	96	26	-71	0.9	180	0.9	55	1	101	1	-176	0.9	-91	1	-134	1
7 G	141	21	-74	0.9	173	0.9	56	1	123	0.9	-164	0.9	-81	0.9	-134	0.9
8 G	80	26	-74	0.9	172	1	53	1	98	1	-	-	-	-	-139	0.9
9 C	2	34	-80	-	176	1	56	1	90	1	-148	0.9	-62	1	-165	1
10 C	20	42	-70	0.9	168	0.9	54	0.9	75	1	-163	0.9	-71	1	-158	1
11 G	7	41	-75	0.9	172	0.9	65	1	82	0.9	-162	0.9	-64	1	-154	1
12 U	13	43	-70	0.9	170	1	61	1	76	1	-159	0.9	-70	0.9	-158	1
13 G	9	38	-74	0.9	171	0.9	64	1	83	1	-166	0.9	-68	1	-161	1
14 U	12	42	-72	1	174	0.9	66	1	78	1	-156	0.9	-65	0.9	-160	1
15 A	7	38	-72	0.9	171	0.9	62	1	82	1	-158	1	-71	1	-161	1
16 G	14	35	-76	1	171	1	64	1	89	1	-	-	-	-	-160	1

Table A8.5: Global base-pair axis parameters for [GEM] structure.

[GEM] duplex DNA strand	Inclination (theta)	Tip	[GEM] duplex RNA strand	Inclination (theta)	Tip
C1	15	8	G16	8	-11
T2	12	-7	A15	11	-8
A3	11	-1	U14	10	-10
dFdC4	2	7	G13	25	-17
A5	15	0.5	U12	7	-5
C6	18	-7	G11	13	-6
G7	20	-8	C10	17	-7
G8	11	-12	C9	22	-13

Table A8.6: Global base-pair axis parameters for [CON] structure.

[CON] duplex DNA strand	Inclination (theta)	Tip	[CON] duplex RNA strand	Inclination (theta)	Tip
C1	17	-4	G16	5	-9
T2	17	-0.4	A15	0.7	-6
A3	17	5	U14	4	-10
C4	20	-5	G13	2	-7
A5	15	3	U12	3	-11
C6	21	1.6	G11	6	-15
G7	19	5	C10	3	-15
G8	21	0.8	C9	11	-14

Table A8.7: Global base-base buckle parameters for [GEM] (left) and [CON] (right) structures.

[GEM] duplex	Buckle (kappa)	[CON] duplex	Buckle (kappa)
C1-G16	7	C1-G16	12
T2-A15	1	T2-A15	16
A3-U14	1	A3-U14	13
dFdC4-G13	-24	C4-G13	18
A5-U12	7	A5-U12	13
C6-G11	5	C6-G11	15
G7-C10	3	G7-C10	16
G8-C9	-11	G8-C9	10

Table A8.8: Global inter-base tilt parameters for [GEM] (left) and [CON] (right) structures.

[GEM] duplex	Buckle (kappa)	[CON] duplex	Buckle (kappa)
C1-T2	-3	C1-T2	0.3
T2-A3	-0.9	T2-A3	0.1
A3-C4	-9	A3-C4	3
dFdC4-A5	13	C4-A5	-5
A5-C6	3	A5-C6	6
C6-G7	2	C6-G7	-2
G7-G8	3	G7-G8	2

CHAPTER 9 ANNEX

Effect of furanose C2'-substitution on structure and stability of G-quadruplex

CONTENTS:

Table A9.1: NMR assignments of **AFtel**.

Table A9.2: Inter-residual 2'F..H8 and H2'..H8 distances for **AFtel** and 1K8P telomeric G-quadruplex. Distances are in Å.

Table A9.3: Inter-residual 2'F..H8-C and H2'..H8-C angles for **AFtel** telomeric structure and 1KP8 telomeric G-quadruplex. Angles are in degrees.

Table A9.4: Inter-residual H2''..O4' distances for **AFtel** telomeric structure and 1K8P telomeric G-quadruplex. Distances are in Å.

Table A9.5: Inter-residual C-H2''..O4' angles for **AFtel** telomeric structure and 1KP8 telomeric G-quadruplex. Angles are in degrees.

Table A9.1: Assignment of the proton and fluorine resonances of 2'-F-araG modified telomeric parallel G-quadruplex. Concentration: 0.7 mM 10 mM potassium phosphate and 5 mM KCl. pH 7.0. T=25 °C).

Residue	NH	NH ₂ (2)	NH ₂ (1)	H6/H8	H2/H5/M	H1'	H2'	H2''	F	H3'
1 THY	n.o	-	-	7.44	1.83	6.27	2.02	2.45	-	4.81
2 ADE	-	n.o	n.o	8.27	8.16	6.15	2.84	3.04	-	5.19
3 GUA	11.9	n.o	n.o	8.25	-	6.37	3.06	3.30	-	5.32
4 GUA	11.6	6.6	9.5	7.95	-	6.35	2.88	3.06	-	5.27
5 GUA	11.3	n.o	8.1	7.94	-	6.51	2.77	2.77	-	5.23
6 THY	n.o	-	-	7.81	2.05	6.53	2.59	2.71	-	5.07
7 THY	n.o	-	-	7.93	2.23	6.55	2.51	2.74	-	5.09
8 ADE	-	n.o	n.o	8.77	8.63	6.84	3.22	3.22	-	5.42
9 2'-FaraG	11.7	n.o	n.o	8.06	-	6.31	-	5.53	-118	5.19
10 GUA	11.6	6.2	9.6	7.69	-	6.25	2.74	2.95	-	5.26
11 GUA	11.4	n.o	7.7	8.00	-	6.47		2.88	-	5.17
12 THY	n.o	-	-	7.37	1.79	6.16		3.04	-	4.66

n.o. not observed

Table A9.2: Inter-residual 2'F..H8 and H2'..H8 distances for **AFtel** and 1K8P telomeric G-quadruplex. Distances are in Å.

Nucleotide	2'F-araG Inter F ..H8 /H2' .. H8 i i+1	X-ray Inter H2' ..H8 i i+1
G3-G4 i i+1	4.1 ± 0.6	3.3
G4-G5 i i+1	4.3 ± 0.2	2.0
G9-G10 i i+1	2.7 ± 0.2	3.8
G10-G11 i i+1	3.6 ± 0.4	3.4

Table A9.3: Inter-residual 2'F..H8-C and H2'..H8-C angles for **AFtel** telomeric structure and 1K8P telomeric G-quadruplex. Angles are in degrees.

Nucleotide	2'F-araG Inter F .. H8C8 /H2' .. H8-C i i+1	X-ray Inter H2' ..H8-C i i+1
G3-G4 i i+1	117	128
G4-G5 i i+1	112	116
G9-G10 i i+1	137	124
G10-G11 i i+1	121	121

Table A9.4: Inter-residual H2''..O4' distances for **AFtel** telomeric structure and 1K8P telomeric G-quadruplex. Distances are in Å.

Nucleotide	2'F-araG Inter H2'' ..O4' i i+1	X-ray Inter H2'' ..O4' i i+1
G3-G4 i i+1	4.1 ± 0.6	3.8
G4-G5 i i+1	4.1 ± 0.6	3.0
G9-G10 i i+1	2.9 ± 0.2	4.4
G10-G11 i i+1	3.7 ± 0.4	4.2

Table A9.5: Inter-residual C-H2''..O4' angles for **AFtel** telomeric structure and 1K8P telomeric G-quadruplex. Angles are in degrees.

Nucleotide	2'F-araG Inter C-H2'' ..O4' i i+1	X-ray Inter C-H2'' .. O4' i i+1
G3-G4 i i+1	112	133
G4-G5 i i+1	108	144
G9-G10 i i+1	152	127
G10-G11 i i+1	132	127

Acrónimos y abreviaturas

ANA: Arabinose

araN: Arabinonucleotide

BCP: Bond Critical Points

CeNA: Ciclohexene

COSY: COrelated Spectroscopy

DFdC: 2'-deoxy-2'-difluorodeoxycytidine

dN: Deoxynucleotide

DNA: Deoxyribonucleic acid

DQF-COSY: Double-Quantum-Filtered COSY

ESI-MS: Electrospray Ionization-Mass Spectrometry

2'F-ANA: 2'-deoxy-2'-fluoro-arabinose

2'F-araN: 2'-deoxy-2'-fluoro-arabinonucleotide

2'F-rN: 2'-deoxy-2'-fluoro-ribonucleotide

2'F-RNA: 2'-deoxy-2'-fluoro-ribonucleic acid

HETCOR: HETeronuclear Correlation

HOESY: Heteronuclear Overhauser Enhancement Spectroscopy

LNA: Locked nucleic acid

MALDI-TOF: Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight

MF: Morpholino

MOE: 2'-O-methoxy-ethyl RNA

mRNA: Messenger RNA

NOE: Nuclear Overhauser Enhancement

NOESY: Nuclear Overhauser Effect Spectroscopy

OMe: 2'-O-methyl RNA

PNA: Peptide nucleic acids

PS: Phosphorothioate

RESP: Restrained ElectroStatic Potential

RISC: RNA-Induced Silencing Complex

rN: Ribonucleotide

RNA: Ribonucleic acids

RNase H: Ribonuclease H

RMN/NMR: Nuclear Magnetic Resonance

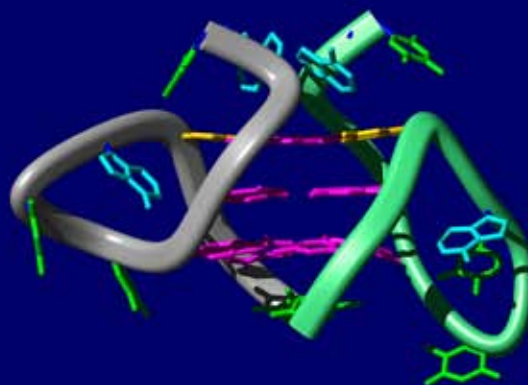
RMSD: Root Mean Square Deviation

ROESY: Rotating frame Overhauser Effect Spectroscopy

siRNA: Small interference RNA

tcDNA: Tricyclo-DNA

TOCSY: TOfal Correlation Spectroscopy



Instituto de Química Física Rocasolano
CSIC
Madrid, junio 2013